

METHOD FOR PRODUCING HEMIN PROTEINS USING PLANT CELLS,  
RESULTING PROTEINS AND PRODUCTS CONTAINING SAME

The invention relates to a method for producing hemin proteins using plant cells, and in particular the hemin proteins capable of reversibly binding oxygen, for example hemoglobin and its derivatives, and myoglobin. It relates, in addition, to the proteins obtained using the method. The invention also relates to the genetically transformed cells and plants capable of producing these proteins, and to the nucleic acid constructs involved in the transformation. In addition, the invention relates to products, for example pharmaceutical or cosmetic products, containing these hemin proteins.

Hemin proteins are complex molecules composed of one or more polypeptide chain(s) in association with one or more ferroporphyrin nucleus or nuclei. These nuclei are composed of four pyrrole rings, juxtaposed in a closed structure and linked by methene bridges, and containing an iron atom at the center of the molecule. Hemin proteins differ from one another in the nature and the number of the polypeptide chains and in the nature of the side chains carried by the eight  $\beta$  atoms of the pyrrole rings. An example of a ferroporphyrin nucleus is iron-containing protoporphyrin IX, also known by the name "protoheme" or simply "heme" (Figure 1).

The hemin protein family comprises numerous substances which are important from the biological point of view in animals and in plants, particularly hemoglobin, myoglobin, cytochromes, peroxidases and catalases.

Hemoglobin is the main constituent of the red blood cells. Its essential function is to bind, transport and deliver the quantity of oxygen necessary for normal tissue function.

The hemoglobin molecule is composed of two types of globin chains or subunits, called  $\alpha$  and  $\beta$  (of

*Moved to page 1 of 4 pages*

- 2 -

(141 and 146 amino acids respectively), and linked in pairs to form a  $\alpha_2\beta_2$  tetramer. Each of these subunits contains, solidly attached in a hydrophobic sac, a heme molecule (that is to say protoporphyrin IX) containing, at the center, a divalent iron atom ( $Fe^{2+}$ ) to which a molecule of oxygen reversibly binds. Each tetrameric hemoglobin molecule therefore contains 4 iron atoms and 4 oxygen molecules which it binds during its passage through the lungs. The molecular mass of the tetramer is 64,650 D. In man, the  $\alpha$  and  $\beta$  chains are synthesized from two types of genes situated on chromosomes 16 and 11 respectively.

The term beta, or "nonalpha", type chains covers not only the beta chains, but also the chains called epsilon, gamma or delta.

Normally, in adults, more than 95% of the hemoglobin consists of alpha<sub>2</sub> beta<sub>2</sub> tetramer, that is to say the association of two heterologous alpha-beta dimers, associated with the catalytic complex, heme. 2% to 3% of a hemoglobin consisting of alpha<sub>2</sub> delta<sub>2</sub> tetramers, and traces of fetal hemoglobin alpha<sub>2</sub> gamma<sub>2</sub> exist.

The tetrameric human hemoglobin molecule exists in two quaternary forms or structures depending on whether oxygen is bound or not to the iron atoms. In the presence of oxygen, hemoglobin is said to be in an R (for relaxed) state and its affinity for oxygen is high. In the absence of oxygen, hemoglobin is said to be in a T (for tense) state and its affinity for oxygen is 100 times lower (Perutz, 1970). The resultant affinity is linked to the equilibrium between the concentrations of R and T forms. The higher the concentration of hemoglobin in the T form at any level of oxygenation, the lower this affinity. The affinity of hemoglobin for oxygen is regulated by the cofactor - 2,3-diphosphoglycerate (DPG), a small molecule derived from the metabolism of glucose and which binds to the  $\beta$  chains of tetrameric hemoglobin, reducing its affinity for oxygen.

(The increase in the risks of infection by products derived from human blood (hepatitis, HIV) makes the development of an artificial oxygen carrier as substitute for blood transfusion necessary.

Techniques using recombinant DNAs have been proposed for producing the protein chains of globin.

The aim of the first techniques developed was essentially to cause the alpha and beta chains to be synthesized in E. coli separately (Nagai and Thogen-Sen, 1987), involving cumbersome methods for separate expression of each of the chains. These methods could hardly be exploited on an industrial scale.

More recently, the expression of soluble and functional recombinant hemoglobin has been developed in E. coli (Hoffman et al., 1990, P.N.A.S., 87, 8521-8525) and Saccharomyces cerevisiae (Wagenbach et al., 1991, Biotechnology, 9, 57-61). Each of these systems has advantages and disadvantages. Indeed, the highest expression levels are obtained in E. coli which has, nevertheless, the disadvantage of producing endotoxins and of not cleaving the NH<sub>2</sub> terminal methionines contrary to Saccharomyces cerevisiae. In the yeast, the yields of synthesis of hemoglobin are low (3 to 5%), compared with the 10-15% obtained in E. coli. This currently limits the use of yeast in the context of an industrial development plan.

The use of animal cells in culture or of transgenic animals as hosts for production has also been achieved (Swanson et al., Bio/Technology, May 1992, 10, page 55). It appears that these techniques cannot currently be exploited because of low expression levels and the risks of contaminations by viruses and by prions.

The technical problem which the present invention proposes to solve is to produce hemin proteins, and in particular hemoglobin and its derivatives, in a large quantity at low costs, without the risk of viral or subviral contaminations. The inventors have provided a solution to this problem by

using plant cells as host for the transformation and the production.

Various teams have already taken an interest in the production of mammalian recombinant proteins in plant cells or in transgenic plants. For example, the specific expression, in rapeseed, of leu-enkephalin has been obtained with expression levels of about 0.1% (Vanderkerckhove et al., Biotechnology, 1989, 7, 929-932).

In 1990, Sijmons et al., (Biotechnology, 1990, 8, 217-221) transferred the gene for human serum albumin into tobacco and potato cells. Regardless of the origin of the signal peptides (human or plant), human serum albumin levels of the order of 0.02% of the total proteins were obtained in the potato leaves, stems and tubers.

Other mammalian recombinant proteins have also been produced in plants: hepatitis B surface antigen (Mason et al., P.N.A.S., 1992, 89, 11745-11749); human interferon (Edelbaum J. of Interferon Res., 1992, 12, 449-453); a mouse antibody to Streptococcus mutans, an agent for dental caries (Hiatt and Ma, FEBS, 1992, 307, 71-75); an anti-Herpes antibody (Russel, 1994) and hirudin (Moloney, 1994). *End page 3*

All these research studies show that the production of mammalian recombinant proteins in plant cells is possible and that the mechanisms of protein synthesis from the DNA sequences are similar in animal cells and plant cells.

On the other hand, little information is available on the subject of the iron-containing porphyrins in plants, particularly on their structures, their synthesis pathways and the assembly of the porphyrin nuclei and the protein chains to form the hemin proteins. The production of recombinant molecules having the capacity to reversibly bind oxygen, and requiring the assembly, in the cell, of heterologous proteins and of endogenous plant porphyrins has never been described.

( The invention relates to a method for producing recombinant hemin proteins using plant cells. According to the method of the invention, the plant cell is genetically modified so as to be able to express the protein component of a hemin protein. The porphyrin nucleus is produced by the cell endogenously, the assembling of the protein and porphyrin components taking place spontaneously by virtue of their high affinity for each other.

More particularly, the invention relates to a method for producing hemin proteins comprising the following steps:

- i) introducing, into plant cells, one or more nucleic acid molecule(s) each of which comprises at least one sequence encoding a protein component of a hemin protein of animal origin or a variant or a portion of this protein component, and optionally a sequence encoding a selection agent;
  - ii) selecting the cells which have integrated the nucleic acid encoding the protein component;
  - iii) propagating the transformed cells, either in culture or by regenerating whole transgenic or chimeric plants;
  - iv) recovering, and optionally purifying, a hemin protein comprising a complex of the protein or proteins encoded by the abovementioned nucleic acid with at least one iron-containing porphyrin nucleus, or a plurality of these complexes.

The invention preferably relates to a method for producing hemin proteins comprising the following steps: ) end page 4

- i) introducing, into plant cells, one or more nucleic acid molecule(s) each of which comprises at least one sequence encoding a protein component of a hemin protein of animal origin preferably capable of reversibly binding oxygen or a variant or a portion of this protein component, and optionally a sequence encoding a selection agent;

- (ii) selecting the cells which contain the nucleic acid encoding the protein component of the hemin protein;
- (iii) optionally, propagating the transformed cells, either in culture or by regenerating whole transgenic or chimeric plants;
- (iv) recovering, and optionally purifying, a hemin protein comprising a complex consisting of the protein or proteins encoded by the abovementioned nucleic acid and at least one iron-containing porphyrin nucleus, or a plurality of these complexes.

In the context of the present invention, the term "hemin protein" means any protein having an iron-containing porphyrin nucleus as prosthetic group, and in particular protoporphyrin IX as exists in human myoglobin and hemoglobin (Figure 1). The porphyrin nucleus may also be derivatives of heme from those of human heme. The side chains are preferably hydrophobic.

The hemin proteins of the invention include in particular the hemin proteins having, as main function, the reversible binding of oxygen, that is to say myoglobin and hemoglobin, as well as the cytochromes whose role is to transport electrons. The derivatives of these proteins conserving these functions are also included in the invention.

According to a preferred variant, the hemin protein of the invention is hemoglobin or a hemoglobin-type protein. In the context of the invention, the term "hemoglobin-type protein" includes all the hemin proteins having at the same time:

- i) one or more  $\alpha$ - and/or  $\beta$ -globin chain(s) or variants of these polypeptides, and
- ii) one or more molecules of iron-containing protoporphyrin IX, or of protoporphyrins differing from protoporphyrin IX in the nature of the side chains,
- iii) having a capacity to reversibly bind oxygen, preferably with an affinity of between 10 and 50 mm Hg at 37°C, pH 7.4. More particularly, the affinity is between 20 and 30 mm Hg, by way of example, the  $P_{50}$  of total blood at pH 7.2 is of 26  $\pm$  2 mm Hg.) End page 5

In the text which follows, the term "hemoglobin-type molecule" will be used synonymously with the term "hemoglobin derivative".

In this context, a "variant" of a protein component, and particularly of  $\alpha$ - or  $\beta$ -globin, means an amino acid sequence which distinguishes itself in relation to the natural sequence by one or more amino acid substitution(s), deletion(s) or insertion(s). In general, the variant exhibits at least 90%, and preferably at least 95%, homology or identity with the natural sequence. In the context of the present invention, the percentage homology between two amino acid sequences is calculated as being the number of identical amino acids plus the number of similar amino acids in the alignment of the two sequences, divided by the length of the sequences between two given positions. If, between the two given positions, the two sequences do not have the same length, the percentage homology is the number of identical and similar amino acids, divided by the length of the longest sequence. The amino acids considered to be "similar" are known in the art, see for example R.F. Feng, M.S. Jobson and R.F. Doolittle; J. Mol. Evol.; 1985; 21; 112-115. They are normally considered to be those which, within a permutation matrix, have a positive coefficient of substitution.

The term "variant" also includes fragments of polypeptide chains, for example of  $\alpha$ - or  $\beta$ -globin, normally having a length of at least 90% of the parent molecule. The variants can also be made longer than the parent molecule by adding nonfunctional sequences. Preferably, the variants conserve the biological and immunological properties of the parent molecule.

The first stage of the method of the invention consists in introducing, into plant cells, one or more nucleic acid molecule(s) comprising at least one sequence encoding a protein component of a mammalian hemin protein, or a variant of this component. )

(When the hemin protein is a single-chain protein, for example myoglobin or cytochrome, the nucleic acid introduced into the plant cells normally comprises a copy of the sequence encoding this protein.

On the other hand, when it is an oligomeric or a multimeric protein, such as hemoglobin or hemoglobin-type molecules, the sequences encoding the various protein units are introduced into the plant cell, either within the same nucleic acid molecule, or within separate nucleic acid molecules. Preferably, for the production of hemoglobin and its derivatives, the sequences encoding  $\alpha$ - and  $\beta$ -globin, or their variants, are within the same vector, called co-expression vector. The vector may comprise one or more copy(ies) of each coding sequence.

Alternatively, the sequences encoding  $\alpha$ - and  $\beta$ -globin, or their variants, may be present on separate nucleic acid molecules. According to this variant, the two molecules may be introduced into the same plant cell, provided that an appropriate selection system is available. Another technique consists in introducing one of the molecules into a first plant cell, and the other into a second plant cell. Each of the transformed cells is then regenerated into a whole plant, it then being possible for the plants thus obtained to be crossed in order to give a progeny capable of producing both the  $\alpha$  and  $\beta$  chains. This approach can be used to optimize the yield of hemoglobin.

The nucleic acid molecules introduced into the plant cell during the first stage of the method are also part of the invention. Generally, these nucleic acids comprise:

- i) one or more sequence(s) encoding a protein component of an animal hemin protein, and
- ii) one or more sequence(s) encoding a targeting signal of plant origin, and/or sequences for regulation of transcription which are recognized by a plant cell.

More particularly, the nucleic acid of the invention comprises: )

- (i) one or more sequence(s) encoding a protein component of an animal hemin protein, the said protein having the capacity to reversibly bind oxygen, and  
ii) sequences for regulation of transcription which are recognized by a plant cell, comprising a promoter and sequences for regulation of termination, and  
iii) one or more sequence(s) encoding a targeting signal of plant origin.

Preferably, the sequences encoding the protein component encode animal  $\alpha$ - or  $\beta$ -globin, for example of human or bovine origin, or the variants thereof. In this manner, the properties of the molecule, and in particular the affinity for oxygen and the stability, can be optimized.

Among these modifications, it is possible, for example, to introduce into one or into both of the  $\alpha$ - and  $\beta$ -globin chains, by site-directed mutagenesis, one or two <sup>start page 4</sup>sequence difference(s) in order to reduce the affinity for oxygen. These mutations may be chosen from examples of natural mutations (see Table I), or from the mutations indicated by examination of the three-dimensional model of natural hemoglobin A. )

end page 7

Moved to page 8  
of the Subst. Rule  
Specification

(Table I: Some mutated human hemoglobins  
(Int. Hemoglobin Center, 1995)

Abnormal hemoglobin	Normal residues and positions	Replacement
$\alpha$ chain		
I	16 Lys	Glu
G <sub>Honolulu</sub>	30 Glu	Gln
Norfolk	57 Gly	Asp
M <sub>Boston</sub>	58 His	Tyr
G <sub>Philadelphia</sub>	68 Asn	Lys
O <sub>Indonesia</sub>	116 Glu	Lys
$\beta$ chain		
C	6 Glu	Lys
S	6 Glu	Val
G <sub>San José</sub>	7 Glu	Gly
E	26 Glu	Lys
M <sub>Saskatoon</sub>	63 His	Tyr
Zurich	63 His	Arg
M <sub>Milwaukee</sub>	67 Val	Glu
D <sub>Punjab</sub>	121 Glu	Gln
Mequon	41 Phe	Tyr
Providence	82 Lys	Asp

In a very advantageous manner, the mutants whose functional properties correspond to the physiological conditions for oxygen transport will be used: reversible binding, cooperativity and low speed of autooxidation. Among the mutants, there will be preferably used the double mutants  $\alpha_2\beta_2$ F41Y, K82D (that is to say a mutant whose  $\beta$  chain comprises the following modifications: Phe-41 is replaced by Tyr, and Lys-82 is replaced by Asp) or  $\alpha_2\beta_2$ F41Y, K66T (that is to say a mutant whose  $\beta$  chain comprises the following modifications: Phe-41 is replaced by Tyr, and Lys-66 is replaced by Thr) which correspond to these functional characteristics.)

and pages

MOVED TO PAGE 8 OF  
THE SUBSTITUTE SPECIFICATION

(The modification of the  $\alpha$  and  $\beta$  chains may also be carried out in order to stabilize the molecule, that is to say to avoid the dissociation of the tetramer into small-sized dimers which are rapidly filtered by the kidneys and which limit the intravascular life of hemoglobin. Covalent bridging, with the aid of phosphate or diaspirin, has been demonstrated as being an effective technique for stabilizing the tetramer (Benesch and Kwong, 1994). The same result can be obtained through modifications of the amino acid chain. The  $\alpha$  subunits are produced in an alpha-alpha dimeric form linked by a glycyl residue. In this form, they conserve their capacity to correctly assemble onto the beta partner subunits and onto heme in order to form a soluble hemoglobin. This hemoglobin can no longer dissociate into dimers because the tetrameric structure is stabilized by a covalent bond (peptide bond) between the alpha-beta dimers. This technique makes it possible to increase the intravascular half-life of the molecule.

Among the variants, it is also possible to use a hybrid protein composed of a portion of the alpha chain and a portion of the beta chain.

According to a preferred variant of the invention, the nucleic acid comprises, in addition to the sequences encoding  $\alpha$ - or  $\beta$ -globin, sequences encoding targeting signals. Preferably, these signals are chloroplast or mitochondrial targeting signals. The expression and/or accumulation of the recombinant proteins in these organelles is particularly preferred because of the availability of endogenous iron-containing porphyrins which are found here. The yield of hemin proteins is therefore increased. In addition, the targeting of the proteins toward the chloroplasts and the mitochondria avoids glycosylation of the protein, which may be advantageous since the natural hemoglobin molecule is not glycosylated.) end page 9

Start page 10 (As an example of chloroplast targeting signals, there may be mentioned the sequence encoding the )  
moved to page 10 of the substitute specification

moved to page 10 of the substrate specification

- 12 -

(transit peptide of the precursor of the small subunit of ribulose 1,5-diphosphate carboxylase of Pisum sativum (see examples). As mitochondrial targeting signals, there may be mentioned the sequence encoding the transit peptide of the precursor of the beta subunit of mitochondrial ATP-aseF1 of Nicotiana plumbaginifolia (see examples).

These transit peptides, as well as the N-terminal methionine, are normally cleaved in the chloroplasts or the mitochondria. The expression of the proteins in the plastids therefore also has the advantage of producing a molecule lacking N-terminal methionine as natural molecule.

According to another variant, the targeting sequences may be sequences encoding an N-terminal signal peptide ("prepeptide"), optionally in association with a signal responsible for retaining the protein in the endoplasmic reticulum (KDEL-type signal), or a vacuolar targeting signal or "propeptide". The presence of the N-terminal signal peptide or prepeptide allows the penetration of the nascent [protein into the endoplasmic reticulum where a certain] amount of post-translational processing occurs, particularly the cleaving of the signal peptide, the N-glycosylations, if the protein in question has N-glycosylation sites, and the formation of disulfide bridges. Among these various signals, the prepeptide responsible for the targeting of the protein into the endoplasmic reticulum, is dominant. It is normally a hydrophobic N-terminal signal peptide having between 10 and 40 amino acids and being of animal or plant origin. Preferably, it is a prepeptide of plant origin, for example that of sporamine, barley lectin, plant extensin,  $\alpha$ -mating factor, pathogenesis protein 1 or 2.

Normally, the signal peptide is cleaved by a peptidase signal upon the co-translational introduction of the nascent polypeptide into the lumen of the RER. The mature protein no longer contains this N-terminal extension.)

(The targeting sequences can, besides the prepeptide, also comprise an endoplasmic retention signal, consisting of the KDEL, SEKDEL or HEKDEL peptides. These signals normally exist at the C-terminal end of the protein and remain on the mature protein. The presence of this signal tends to increase the recombinant protein yields.) *end page 10*

(The targeting signals may, besides the *start* prepeptide, also comprise a vacuolar targeting signal *page 11* or "propeptide". In the presence of such a signal, after passing into the RER, the protein is targeted toward the vacuoles of the aqueous tissues, the leaves for example, as well as to the protein bodies of the storage tissues, for example the seeds, tubers and roots. The targeting of the protein toward the protein bodies of the seed is particularly advantageous because of the capacity of the seed to accumulate proteins, up to 40% of the proteins relative to the dry matter, in cellular organelles derived from the vacuoles, called protein bodies and because of the possibility of stocking, for several years, the seeds containing the recombinant proteins in the dehydrated state.

As propeptide, it is possible to use a signal of animal or plant origin, the plant signals being particularly preferred, for example prosporamine. The propeptide may be N-terminal ("N-terminal targeting peptide" or NTTP), or C terminal (CTTP) Since the propeptides are normally cleaved upon entry of the protein into the vacuole, it is not present in the mature protein.

The use of the signal peptide or prepeptide can lead to the glycosylation of the protein. Normally, globin has no N-glycosylation sites, but these may be introduced by mutagenesis. The  $\alpha$  and  $\beta$  chains can also have O-glycosylation sites.

In the absence of any targeting signal, the protein is expressed in the cytoplasm.

The nucleic acid introduced into the plant cell may also comprise sequences for regulation of)

moved to page 12 of the  
Substitution Specification

Endorse =  
Endorse =

- 14 -

(transcription which are recognized by the plant cell. The nucleic acid is in this case a "chimeric gene". The regulatory sequences comprise one or more promoter(s) of plant or viral origin or obtained from Agrobacterium tumefaciens. They may be constitutive promoters, for example the CaMV 35S, the double 35S, the Nos or OCS promoters, or promoters specific for certain tissues such as the grain or specific for certain phases of development of the plant. As promoters specific for seeds, there may be mentioned the promoter of the gene for napin and for the acyl carrier protein (ACP) (EP-A-0,255,378), as well as the promoters of the AT2S genes of Arabidopsis thaliana, that is to say the PAT2S1, PAT2S2, PAT2S3 and PAT2S4 promoters (Krebbers et al., Plant Physiol., 1988, vol. 87, pages 859-866). It is) (particular) <sup>Start page (2)</sup> preferable to use the cruciferin or phaseolin promoter or pGEA1 and pGEA6 of Arabidopsis, promoters of genes of the "Em, Early Methionine labelled protein" type, which is strongly expressed during the phases of drying of the seed.

It is possible to envisage using "enhancers" to improve the efficiency of expression. When the transformation occurs directly in the chloroplast and mitochondrial genomes, gene promoters specific for these compartments can be used.

The sequences for regulation of transcription normally comprise sequences for termination of transcription which are of plant or of viral origin, for example 35S, or of bacterial origin (Agrobacterium).

When the transforming nucleic acid does not comprise regulatory sequences, it is preferable to add onto each end of the nucleic acid a DNA sequence homologous to the genomic sequences which are adjacent to a specific insertion site in the genome. This allows the integration of the construct by homologous recombination, at a site where endogenous regulatory sequences can control the expression of the heterologous sequences. )

moved to page 12 of the  
Substitution Specification

Moved to page 12 of the  
Substitute Specification

Start page  
B

Moved to page 13 of the  
Substitute Specification

- 15 -

(The nucleic acids of the invention may also comprise one or more intron(s), preferably of plant origin. These introns, which are obtained from a plant gene, are introduced artificially so as to increase the efficiency of expression of the heterologous sequence.

Indeed, it has been demonstrated, particularly in monocotyledonous plants, that the insertion of an intron into the untranslated 5' portion of a gene, that is to say between the site of initiation of transcription and the site of initiation of translation, leads to an improvement in the stability of the messenger, and consequently, to a better expression. The intron(s) used in this manner are obtained preferably from a monocotyledonous plant such as maize. This is preferably, but not necessarily, the first intron of the gene.

The nucleic acid sequence encoding  $\alpha$ - and  $\beta$ -globin and its variants is normally cDNA. Appropriate sequences are illustrated in Figures 2 and 3, any degenerate sequence can also be used as well as the sequences of the variants as defined above.) End page 12

~~~~~(The introduction of a nucleic acid molecule(s) into the plant cell can be carried out in a stable manner either by transformation of the nuclear genome, or by transformation of the chloroplast genome of the plant cell, or by transformation of the mitochondrial genome.

For the transformation of the nuclear genome, conventional techniques may be used. All known means for introducing foreign DNA into plant cells may be used, for example Agrobacterium, electroporation, protoplast fusion, particle gun bombardment, or penetration of DNA into cells such as pollen, microspore, seed and immature embryo. Viral vectors such as the Gemini viruses or the satellite viruses may also be used as introducing means. Agrobacterium tumefaciens and rhizogenes constitute the preferred means. In this case, the sequence of the invention is introduced into an appropriate vector with all the )

(necessary regulatory sequences such as promoters, terminators and the like, as well as any sequence necessary for selecting the transformants which have integrated the heterologous sequences.

The transformation of the nuclear genome of the plant cell is often carried out using the targeting signals mentioned above and which determine the cellular compartment where the expression and/or accumulation of the protein will occur.

According to another variant of the invention, the introduction of the nucleic acid into the plant cell can be carried out by the transformation of the mitochondrial or chloroplast genomes (see for example Carrer et al., Mol. Gen. Genet., 1993, 241, 49-56).

Techniques for direct transformation of the chloroplasts or the mitochondria are known per se and may comprise the following steps:

- i) introducing transformant DNA by the biolistic technique (Svab et al., P.N.A.S., 1990, 87, 8526-8530);
- ii) integrating the transformant DNA by two homologous recombination events;
- iii) selectively removing copies of the wild-type genome during repeated cell divisions on selective medium.

In order to allow the homologous recombination of the transformant DNA, two DNA fragments homologous to the genomic sequences, for example the <sup>end of p. 13</sup> (rbcL and ORF 512 genes are added to each end of the DNA to be inserted into the genome.

The direct transformation of the chloroplasts or mitochondria has the advantage of substantially increasing the yield of hemoglobin but the N-terminal methionine is retained.

According to another variant of the invention, the heterologous nucleic acid can be introduced into the plant cell by means of a viral vector.

The method of the invention comprises a step of detecting the hemin proteins and in particular hemoglobin and its derivatives. This makes it possible

MOVED TO PAGE 13 OF THE SUBSTANTIAL SPECIFICATION

MOVED TO PAGE 14 OF THE SUBSTANTIAL SPECIFICATION

(to verify if the plant or the plant cell is capable, not only of expressing the heterologous proteins, but also of assembling them correctly with the porphyrin nucleus. For the hemoglobin in a complex environment containing other chromophores or molecules which scatter light, detection by time-resolved optical spectroscopy will be advantageously used. This technique is described in detail in the examples. Other detection techniques consist in using antibodies specific for the alpha or beta globin chains or their variants. The spectrometric and immunological techniques can be used in association with each other. The use of these techniques makes it possible to select the plants which are capable of producing hemoglobin and its derivatives according to the invention.

The method of the invention comprises, in addition, a step of recovering or extracting hemoglobin or its derivatives from plant tissues. The extraction is normally carried out by grinding the tissues, for example leaves or grains, in an appropriate buffer, filtering the ground product, precipitating the proteins in the supernatant, centrifuging and taking up the pellet in an appropriate buffer with dialysis. A partial purification step can also be carried out at this stage by chromatography on an anion-exchange column.

The tetramer of hemoglobin, or of its derivatives, is purified by two successive chromatographies on an ion-exchange resin followed by a step of concentrating and saturating the concentrate with carbon monoxide. These techniques are described in detail in the examples.) end page 14

(When the expression of hemoglobin and of its derivatives takes place under the control of a constitutive promoter, such as the 35S double promoter, an expression level of at least 1% hemoglobin compared with the total proteins may be obtained. The proteins represent about 10% of the dry mass of the leaf and a ton of dry tobacco leaves is harvested per hectare. It )

moved to page 14 of the  
specification  
of the sub-site  
specification

(is therefore possible to obtain of the order of 100 grams of hemoglobin per hectare of tobacco cultivated, assuming that only 10% of the hemoglobin produced is purified.

The method of the invention therefore allows the production of hemoglobin at very low costs with a higher production capacity than that obtained using fermenters of the culture of bacteria or yeast.

Besides the method of transformation, the invention also includes vectors comprising one or more nucleic acid(s) or chimeric gene(s) defined above. As an example of vectors, there may be mentioned binary vectors or plasmids, viral vectors such as gemini viruses or the CaMVs.

The invention also relates to the plant cells transformed with the nucleic acid sequences of the invention. Preferably, they are transformed plant cells capable of producing one or more hemoglobin(s) or derivatives of hemoglobin according to the invention.

They may be plant cell cultures in vitro, for example in liquid medium. Various modes of culture ("batch", "fed batch" or continuous) for this type of cells are currently under study. The "batch" cultures are comparable to those carried out in an Erlenmeyer flask since the medium is not changed, these cells thus have only a limited quantity of nutrient materials. The "fed batch" culture corresponds, for its part, to a "batch" culture with programmed supply of substrate. For a continuous culture, the cells are supplied continuously with nutrient medium. An equal volume of the biomass-medium mixture is removed in order to maintain the volume of the reactor constant. The quantities of plant biomass which can be envisaged with cultures in bioreactors are variable depending on the plant species, the mode of culture and the type of bioreactor. Under certain conditions, biomass densities of about 10 to 30 g of dry weight per liter of culture can be obtained for species such as Nicotiana tabacum, Vinca rosea and Catharanthus roseus. ) end page 15

move to page 15 of the subsequent section

The cells of the invention can also be immobilized, which makes it possible to obtain a constant and prolonged production of hemoglobin. The separation of the hemoglobin and the plant biomass is also facilitated. As immobilization method, there may be mentioned immobilization in alginate or agar beads, inside polyurethane foam, or alternatively inside hollow fibers.

The cells of the invention may also be root cultures. The roots cultivated in vitro, in a liquid medium, are called "Hairy roots", they are roots transformed by the bacterium Agrobacterium rhizogenes.

Instead of producing the hemoglobin of the invention by culturing plant cells, it is possible to regenerate chimeric or transgenic plants from transformed explants, using techniques known per se.

As appropriate plants, there may be mentioned the Angiospermae comprising monocotyledonous and dicotyledonous plants. More particularly, there may be mentioned tobacco, species belonging to botanic families such as leguminous plants (for example beans, peas and the like), cruciferous plants (for example cabbage, radish, rapeseed and the like), Solanaceae (for example tomatoes, potato and the like), Cucurbitaceae (for example melon), Chenopodiaceae (for example beetroot), Umbelliferae (for example carrots, celery and the like). There may also be mentioned cereals such as wheat, maize, barley, triticale and rice, oleaginous plants such as sunflower and soybean. Tobacco, potato, tomato and maize are particularly preferred. For potato, the expression takes place preferably in the tubers.

The invention also relates to the seeds of transgenic plants capable of producing hemoglobin as well as their progeny.

The invention also relates to the hemin proteins which may be obtained by the method of the invention, in particular the hemin proteins capable of )

*Moved to page 16 of the subject specification*

- 2 -  
reversibly binding oxygen, for example the hemoglobins and derivatives thereof.

The hemoglobins of the invention are capable of binding O<sub>2</sub> in a reversible manner with an affinity (P<sub>50</sub>) preferably close to physiological values (37°C, pH 7.40). The affinity of the molecule for O<sub>2</sub> is expressed as P<sub>50</sub>: that is to say the partial pressure of O<sub>2</sub> when hemoglobin or its derivatives is 50% saturated. The P<sub>50</sub> is measured according to the usual techniques, for example by means of an analyzer which measures the percentage O<sub>2</sub> saturation as a function of the O<sub>2</sub> pressure (Kister et al., 1987). Normally, the hemoglobins of the invention have an acceptable autoxidation rate in order to minimize the formation of methemoglobin which is unsuited to the transport of O<sub>2</sub>. This characteristic can be measured by the absorption spectrum.

Preferably, the hemoglobins of the invention are tetramers, preferably alpha<sub>2</sub> beta<sub>2</sub>, beta<sub>4</sub>, or optionally tetramers of chimeric α/β subunits (Dumoulin et al., 1994, Art. Cells, Blood Subst., and Immob. Biotech., 22, 733-738) or multiples of four subunits. The physical size of the complex should be at least that of the tetramer in order to avoid its filtration by the kidneys.

The hemin proteins of the invention can be used in numerous pharmaceutical, cosmetic or industrial applications. The invention relates in particular to pharmaceutical compositions comprising one or more hemin protein(s) according to any one of Claims 15 to 23, in association with a physiologically acceptable excipient.

In the pharmaceutical field, all the conditions requiring an improvement in the transport of oxygen can be treated with the hemoglobins of the invention, these conditions comprising the following:

- acute or chronic hemorrhage,
- states of shock,
- coronary or sylvian angioplasties,

- ( - treatments of solid tumors, sensitization to gamma-therapy,
- preservation of organs before transplant and during transport,
- malignant hemopathies.

The hemoglobins of the invention are normally used in the form of an injection in solutions optionally stabilized as regards the tetrameric form of the complex (for example addition of pyridoxal phosphate or diaspirin) as regards autooxidation. It is also possible to use suspensions of hemoglobin grafted on a support in order to increase the lifetime in the bloodstream. The support may be any conventional support in this domain, for example polysaccharides.

Various aspects of the invention are illustrated in the figures:

- Figure 1: Iron-containing protoporphyrin III (IX), ) end page 17
- ( - Figure 2: cDNA sequence of human  $\alpha$ -globin (423 base pairs), and corresponding protein,
- Figure 3: cDNA sequence of human  $\beta$ -globin (438 base pairs), and corresponding protein,
- Figure 4: Experimental device for flash photolysis. A pulsed laser serves for the photo-dissociation of the Hb ligands:  $(HbCO \rightarrow Hb + CO)$ . A second optical beam, oriented at  $90^\circ$ , detects changes in absorption as a function of time after dissociation.
- Figure 5: Kinetics of bimolecular recombination of CO with hemoglobin in plant extract. The two phases correspond to the two allosteric states of Hb: R (rapid) and T (slow).  
Conditions: 0.1 atm CO, pH 6-6, 25°C, about 50% dissociation, 0-1, 1 and 10  $\mu M$  Hb.
- Figure 6: Kinetics of recombination of CO with hemoglobin as a function of the percentage dissociation (by variation of the laser energy). The kinetics are sensitive to the number of ligands dissociated (1 to 4). At a high level of dissociation, Hb (deoxy or mono-ligand containing) shifts toward the

(slow form "T". At a low laser energy, the tetramers (mainly with three ligands) remain in the rapid form "R".

- Figure 7: Western-blot analysis of the extract of the seeds of the transgenic tobacco T26-22 transformed with the plasmid pBIOC59. The extracts of seeds (75 µg of proteins) of a nontransformed tobacco (1) and of a transgenic tobacco (T26-22) (2), molecular weight markers (3) and HbA (50 ng) (4) are separated by SDS-PAGE 17% electrophoresis under reducing conditions. The Western blotting is carried out under the conditions described in section X.a. The molecular weight markers and the α and β globins are indicated.

- Figure 8: Western-blot analysis of the fractions obtained during partial purification. The proteins in the fractions eluted from Sephadryl S-100 (37 µg), from S-Sepharose (30 µg) which are obtained during purification from mixtures of control seeds [(2) and FE-Control (4) respectively] and of seeds accumulating rHb [(3) and FE-rHb (5) respectively], HbA (50 ng) (1) and molecular weight markers (6) were separated by SDS-PAGE 17% electrophoresis under reducing conditions. The Western blotting is carried out under the conditions described in section X.a. The molecular weight markers and the α and β globins are indicated.

- Figure 9: Kinetics of recombination of CO with the FE-rHb fraction. The kinetics, following flash photolysis, is characteristic of the normal tetrameric Hb. The FE-Control fraction obtained from the control plants gives a signal of amplitude 1 mOD, that is to say about 50 times weaker than that observed for the FE-rHb fraction (48 mOD).

- Figure 10: Kinetics of recombination of CO with the FE-rHb fraction at various laser intensity levels. Similar results are observed for HbA (Figure 6).

- Figure 11: Demonstration of the reversible binding of oxygen to the FE-rHb sample. Since the )

Moved to page 19 of  
The Substitute Specification

- 23 -

oxyhemoglobin samples give only weak signals, we used, for these measurements, the technique of mixing a CO atmosphere and O<sub>2</sub>. After photodissociation of CO, the rapid phase corresponds to the binding of oxygen. The oxygen is then replaced with CO which can again be photodissociated. The figure also shows the kinetics of recombination of the CO of the same sample equilibrated under 1 atm or 0.1 atm CO.

#### EXAMPLES

##### I. CONSTRUCTION OF BASAL EXPRESSION BINARY PLASMIDS ALLOWING THE PRODUCTION OF RECOMBINANT PROTEINS IN TOBACCO LEAVES.

The expression of genes in tobacco leaves requires the following regulatory sequences:

1) the constitutive double 35S promoter (pd35S) of CaMV (cauliflower mosaic virus). It corresponds to a duplication of the transcription-activating sequences situated upstream of the TATA element of the natural 35S promoter (Kay et al., 1987).

2) the sequence for termination of transcription, 35S polyA terminator, which corresponds to the noncoding 3' region of the sequence of the circular (double-stranded DNA cauliflower mosaic virus producing the 35S transcript (Franck et al., 1980)).

The constructions of the various plasmids via the use of recombinant DNA techniques (Sambrook et al., 1989) are derived from pBIOC4. This binary plasmid is derived from pGA492 (An, 1986) which contains, between the right and left borders derived from the plasmid pTiT37 of Agrobacterium tumefaciens, on its transfer DNA, the following sequences:

- the constitutive promoter of the nos gene encoding nopaline synthase (Depicker et al., 1982),
- the coding sequence of the nptII gene encoding neomycin phosphotransferase II (Berg and Berg, 1983) deleted off the region of the first 8 codons)

end page 19

Moved to page 20 of  
The Substitute Specification

( including the ATG methionine initiation codon and fused to the sequence of the first 14 codons of the coding sequence of the nos gene (Depicker et al., 1982), the coding sequence of the nos gene lacking the region of the first 14 codons, the nos terminator (Depicker et al., 1982), a polylinker (HindIII-XbaI-SacI-HpaI-KpnI-ClaI-BglII) preceding the cat gene encoding chloramphenicol acetyltransferase (Close and Rodriguez, 1982) and the termination sequences of gene 6 of the plasmid pTiA6 of Agrobacterium tumefaciens (Liu et al., 1993).

To remove virtually the whole of the coding sequence of the cat gene, the plasmid pGA492 was doubly digested with SacI (restriction site of the polylinker) and with ScaI (restriction site present in the sequence of the cat gene) and then subjected to the action of the enzyme T4 DNA polymerase (New England Biolabs) according to the manufacturer's recommendations. The ligation of the modified plasmid (20 ng) was carried out in a reaction medium of 10 µl containing 1 µl of 10× T4 DNA ligase buffer (Amersham) and 2.5 U of T4 DNA ligase enzyme (Amersham) at 14°C for 16 hours. The E. coli DH5α bacteria previously made competent were transformed (Hanahan, 1983).

The plasmid DNA of the clones obtained, selected on 12 µg/ml tetracycline, was extracted according to the alkaline lysis method (Birnboim and Doly, 1979) and analyzed by enzymatic digestion with restriction enzymes. Next, the HindIII restriction site of the plasmid DNA of the selected clone was modified at an EcoRI) (restriction site with the aid of a phosphorylated HindIII - EcoRI adaptor (Stratagene Cloning Systems). To carry out this modification, 500 ng of plasmid DNA of the selected clone were digested with HindIII, dephosphorylated with the enzyme calf intestinal alkaline phosphatase (Boehringer Mannheim) according to the manufacturer's recommendations and coprecipitated in the presence of 1500 ng of HindIII-EcoRI adaptor DNA, 1/10 volume of 3 M sodium )

Moved to page 20 of the 'Substrate Specific' section

end page 20

Moved to page 21  
of the 'Substrate Specific' section

Moved to page 21 of the substituted heteropolymer specification

- 15 -

acetate pH 4.8 and 2.5 volumes of absolute ethanol at -80°C for 30 min. After centrifugation at 12000 g for 30 min, the precipitated DNA was washed with 70% ethanol, dried, taken up in 8 µl of water, heated at 65°C for 10 min, and then ligated in the presence of 1 µl of 10× T4 DNA ligase buffer (Amersham) and 2.5 U of the enzyme T4 DNA ligase (Amersham) at 14°C for 16 hours. After inactivation of the T4 DNA ligase at 65°C for 10 min, the ligation reaction mixture was digested with EcoRI, purified by electrophoresis on a 0.8% agarose gel, electroeluted (Sambrook et al., 1989), precipitated in the presence of 1/10 volume of 3 M sodium acetate pH 4.8 and 2.5 volumes of absolute ethanol at -80°C for 30 min, centrifuged at 12000 g for 30 min, washed with 70% ethanol and then dried. The E. coli DH5α bacteria previously made competent were transformed (Hanahan, 1983). The plasmid DNA of the clones obtained, selected on 12 µg/ml tetracyclin, was extracted according to the alkaline lysis method (Birnboim and Doly, 1979) and analyzed by enzymatic digestion with HindIII and EcoRI in particular. The resulting binary plasmid, which now possesses only the last 9 codons of the coding sequences of the cat gene and in which the EcoRI site is unique, was called pBIOC4.

a. CONSTRUCTION OF THE EXPRESSION BINARY PLASMID pBIOC21.

The expression cassette, consisting of the pd35S promoter and the 35S polyA terminator, was isolated from the plasmid pJIT163D. The plasmid pJIT163D is derived from the plasmid pJIT163 which is itself derived from the plasmid pJIT60 (Guerineau and Mullineaux, 1993). The plasmid pJIT163 possesses an ATG codon between the HindIII and SalI sites of the polylinker. To eliminate this ATG and to obtain the plasmid pJIT163D, the plasmid DNA pJIT163 was doubly digested with HindIII and SalI, purified by )

move  
to page 21  
of six pages  
specification

electrophoresis on a 0.8% agarose gel, electroeluted (Sambrook et al., 1989), precipitated in the presence of 1/10 volume of 3 M sodium acetate pH 4.8 and 2.5 volumes of absolute ethanol at -80°C for 30 min, centrifuged at 12000 g for 30 min, washed with 70% ethanol, dried, subjected to the action of the Klenow enzyme (New England Biolabs) according to the manufacturer's recommendations, deproteinized by extraction with 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1) and then 1 volume of chloroform:isoamyl alcohol (24:1), precipitated in the presence of 1/10 volume of 3 M sodium acetate pH 4.8 and 2.5 volumes of absolute ethanol at -80°C for 30 min, centrifuged at 12000 g for 30 min, washed with 70% ethanol, dried and finally ligated in the presence of 1 µl of 10× T4 DNA ligase buffer (Amersham) and 2.5 U of T4 DNA ligase enzyme (Amersham) at 14°C for 16 hours. The *E. coli* DH5α bacteria previously made competent, were transformed (Hanahan, 1983). The plasmid DNA of the clones obtained, selected on 50 µg/ml ampicillin, was extracted according to the alkaline lysis method (Birnboim and Doly, 1979) and analyzed by enzymatic digestion with restriction enzymes. To isolate the expression cassette consisting of the pd35S promoter and of the 35S polyA terminator (SacI-XhoI fragment), the plasmid DNA of the pJIT163D clone selected was digested with SacI and XhoI. The SacI-XhoI-fragment, carrying the expression cassette, was purified by electrophoresis on a 0.8% agarose gel, electroeluted (Sambrook et al., 1989), precipitated in the presence of 1/10 volume of 3 M sodium acetate pH 4.8 and 2.5 volumes of absolute ethanol at -80°C for 30 min, centrifuged at 12000 g for 30 min, washed with 70% ethanol, dried and then subjected to the action of Mung Bean nuclease enzyme (New England Biolabs) according to the manufacturer's recommendations. This purified insert (200 ng) was cloned into the plasmid DNA of pBIOC4 (20 ng) digested with EcoRI, treated with the Mung Bean nuclease enzyme and dephosphorylated with )

end page 21

to page 22 of the  
Substrate Specification

purified to

(the enzyme calf intestinal alkaline phosphatase (Boehringer Mannheim) according to the manufacturer's recommendations. The ligation reaction was carried out in 20  $\mu$ l, in the presence of 2  $\mu$ l of 10x T4 DNA ligase buffer (Amersham), 2  $\mu$ l of 50% polyethylene glycol 8000 and 5 U of T4 DNA ligase enzyme (Amersham) at 14°C for 16 hours. The E. coli DH5 $\alpha$  bacteria previously made competent were transformed (Hanahan, 1983). The plasmid DNA of the clones obtained, selected on 12  $\mu$ g/ml tetracyclin was extracted according)(to the alkaline lysis method (Birnboim and Doly, 1979) and analyzed by enzymatic digestion with restriction enzymes. The resulting plasmid was called pBIOC21.

b. CONSTRUCTION OF THE CO-EXPRESSION BINARY PLASMID pBIOC43.

The co-expression binary plasmid will allow expression of two genes in the same binary vector.

The co-expression binary plasmid is derived from pBIOC21. It contains two expression cassettes each consisting of a pd35S promoter and a 35S polyA terminator but differ in the polylinker separating the promoter from the terminator. One of the expression cassettes is that of pBIOC21 already described in paragraph I.a. The other expression cassette was obtained by replacing the HindIII-BamHI-SmaI-EcoRI polylinker of pJIT163D (described in paragraph I.a.) by a HindIII-EcoRI adaptor carrying the PacI, AscI, MluI and HpaI restriction sites. This adaptor was obtained by renaturation of the 2 oligodeoxynucleotides WD11 (5' AGC TGA TTA ATT AAG GCG CGC CAC GCG TTA AC 3') and WD12 (5' AAT TGT TAA CGC GTG GCG CGC CTT AAT TAA TC 3') which are complementary for their 28 terminal 3' nucleotides. One hundred  $\mu$ M of each of these two oligodeoxynucleotides were previously phosphorylated by the action of 10 U of T4 polynucleotide kinase enzyme (New England Biolabs) in a total reaction volume of 10  $\mu$ l containing 1  $\mu$ l of 10x T4 polynucleotide kinase )

(buffer (New England Biolabs) and 3  $\mu$ l of ATP (95 mM). The two reaction mixtures were incubated at 37°C for 1 hour, and then at 65°C for 20 min. They were then combined and their volume was adjusted to 500  $\mu$ l. After extraction with 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1) and 1 volume of chloroform:isoamyl alcohol (24:1), 50  $\mu$ l of 3 M sodium acetate pH 6.0 were added. The reaction mixture was incubated at 80°C for 10 min and then cooled slowly to room temperature. The DNA was then precipitated in the presence of 2.5 volumes of absolute ethanol at -80°C for 30 min, centrifuged at 14000 g at 4°C for 1 hour, washed with 70% ethanol, centrifuged at 14000 g at 4°C for 10 min, dried, taken up in 10  $\mu$ l of H<sub>2</sub>O. The HindIII-EcoRI DNA fragment was then cloned at the HindIII-EcoRI sites of the)(plasmid DNA pJIT163D previously dephosphorylated with the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation reaction was carried out in a reaction volume of 20  $\mu$ l in the presence of 1 U of T4 DNA ligase (Gibco-BRL) for a total DNA concentration of 8.5 nM with a vector/insert molar ratio of 1 and of 4  $\mu$ l of 5x T4 DNA ligase buffer (Gibco-BRL) at 25°C for 16 hours. The E. coli DH5 $\alpha$  bacteria previously made competent were transformed (Hanahan, 1985). The plasmid DNA of the clones obtained, selected on 100  $\mu$ g/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC42. Its validity was verified by sequencing with the aid of the "Sequenase Version 2.0 DNA Sequencing" kit marketed by United States Biochemical (USB) according to the dideoxynucleotides method (Sanger et al., 1977). The reaction conditions follow the manufacturer's recommendations except for the denaturation and hybridization. The reaction medium containing the plasmid DNA (0.5 to 1 pmol), the oligonucleotide primer (2 pmol), 10% DMSO and the 1x)

Moved to page 24 of the  
Subunit specification

- 29 -

(reaction buffer (USB), is incubated at 100°C for 10 min, then suddenly cooled to -80°C in dry ice.

From pBIOC42, the DNA fragment encoding the expression cassette consisting of the pd35S promoter and of the 35S polyA terminator was isolated by double digestion with SacI and XhoI. It was purified by electrophoresis on a 0.75% agarose gel, and then subjected to the action of the "Geneclean II" kit marketed by BIO101 according to the manufacturer's recommendations. Next, this DNA fragment was inserted at the SacI and XhoI sites of the plasmid pBCSK+ marketed by Stratagene and previously dephosphorylated with the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation was carried out in a reaction volume of 20 µl in the presence of 1 U of T4 DNA ligase (Gibco-BRL) for a total DNA concentration of 8.5 nM with a vector/insert molar ratio of 1 and of 4 µl of 5x T4 DNA ligase buffer (Gibco-BRL) at 25°C for 16 hours. The E. coli DH5 $\alpha$  bacteria previously made competent were transformed (Hanahan, 1985). The plasmid DNA of the clones obtained, selected on 30 µg/ml of chloramphenicol, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed) (by enzymatic digestion with restriction enzymes. The resulting plasmid was called pBIOC75.

From pBIOC75, the DNA fragment carrying the expression cassette consisting of the pd35S promoter and the 35S polyA terminator was isolated by digestion with KpnI. It was purified by electrophoresis on a 0.75% agarose gel, and then subjected to the action of the "Geneclean II" kit marketed by BIO101 according to the manufacturer's recommendations. Next, this DNA fragment was ligated to the plasmid DNA of pBIOC21 digested with KpnI and dephosphorylated with the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation was carried out in a reaction volume of 20 µl in the presence of 1 U of T4 DNA ligase )

Moved to page 25 of  
The Subunit specification

Moved to page 25 of the  
Substrate Specification

- 30 -

(Gibco-BRL) for a total DNA concentration of 8.5 nM with a vector/insert molar ratio of 1 and of 4 µl of 5x T4 DNA ligase buffer (Gibco-BRL) at 25°C for 16 hours. The *E. coli* DH5α bacteria previously made competent were transformed (Hanahan, 1985). The plasmid DNA of the clones obtained, selected on 12 µg/ml of tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion with restriction enzymes. The resulting plasmid was called pBIOC43.

## II. CONSTRUCTION OF THE CHIMERIC GENES ENCODING THE α AND β GLOBIN CHAINS ALLOWING EXPRESSION OF RECOMBINANT HUMAN HEMOGLOBIN IN THE CYTOPLASM OF TOBACCO LEAVES.

The plasmid alpha1pJW101 contains the cDNA for the α globin chain, cloned into the plasmid pMB9 as described by Wilson et al. (1978).

The M13mp10 phage cIIIX beta contains the cDNA for the β globin chain, cloned into the M13 mp10 phage as described by Nagai et al. (1985). In this construction, the cDNA encoding β globin was inserted in 3' of the coding sequence for the cII protein of the lambda phage, followed by that encoding the FX tetrapeptide, forming a fusion gene in which the initiator ATG codon of β globin has been deleted.) end phage 25

### a. CONSTRUCTION OF THE PLASMID pBIOC44 CONTAINING THE cDNA ENCODING α GLOBIN FOR CYTOPLASMIC TARGETING.

To obtain cytoplasmic targeting of the α globin chain, the initiator methionine codon of the α globin chain was maintained.

The cDNA encoding the cytoplasmic targeting α globin chain was obtained in three stages. The first two stages made it possible to suppress the internal HindIII site (substitution of a T for a C) at position 276 of the coding sequence whereas the third stage )

Moved to page 26  
of the substrate  
specification

Moved to page 26 of the  
Substrate Specificity

- 31 -

Combines the 2 cDNA fragments encoding the recombinant α globin chain.

The first stage consisted in the amplification of the first 95 codons of the mature α globin chain on the plasmid alpha1pJW101 with the aid of the 2 oligodeoxynucleotides, WD13 (5' tacaagcttaaca ATG GTG CTG TCT CCg GCC GAC 3') and AD27 (5' CGG GTC CAC CCG GAG CTT GTG 3'). The WD13 primer provides the HindIII restriction site, the sequence aaca favoring the initiation of translation (Joshi, 1987) and preceding the initiator ATG codon followed by the first 6 codons of the mature α globin chain of which the fourth (CCT) is substituted for CCg (silent mutation) in order to create the EagI restriction site. The AD27 primer allows the suppression of the HindIII restriction site by substitution of nucleotide T for C (position 276 of the coding sequence). The PCR amplification was carried out in 100 µl of reaction medium containing 10 µl of 10× Taq DNA polymerase buffer (100 mM Tris-HCl pH 8.4, 500 mM KCl and 20 mM MgCl<sub>2</sub>), 16 µl of the dNTP mixture (1.25 mM dATP, 1.25 mM dCTP, 1.25 mM dGTP and 1.25 mM dTTP), 10 µl of each of the primers described above at 10 µl, 10 µl of template DNA (alpha1pJW101) at 1 ng/µl and 0.5 µl of Taq DNA polymerase at 5 U/µl (Perkin Elmer). Thirty cycles each comprising 30 sec of denaturation at 97°C, 1 min hybridization at 55°C and 2 min extension at 72°C were carried out in the Appligène "Crocodile II" apparatus. The amplified DNA fragments were then purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit marketed by BIO101 according to the manufacturer's recommendations. The purified amplified DNA fragments are taken up in 20 µl.

The second stage consisted in the amplification of the last 54 codons of the mature α globin chain on the plasmid alpha1pJW101 with the aid of the 2 oligodeoxynucleotides, AD26 (5' CAC AAG CTC CGG GTG GAC CCG 3') and WD14 (5' gcgaaattc TCA ACG GTA TTT GGA GGT CAG CAC 3'). The WD14 primer provides the EcoRI)

Moved to p. 27  
of the substrate specificity

and page 28

(restriction site situated just after the stop codon. The AD26 primer allows the suppression of the HindIII restriction site by substitution of nucleotide T for C (position 276 of the coding sequence). The PCR amplification was carried out as described in the first stage. The treatment of the amplified DNA fragments was carried out as described in the first stage.

The third stage was the PCR amplification of the complete cDNA encoding the  $\alpha$  globin chain (142 codons including the initiator ATG). The two types of DNA fragments amplified in the first and second stages served as template DNA and the two primers used were WD13 and WD14. The PCR amplification was carried out as described in the first stage except that the hybridization temperature of the cycle is 60°C. The amplified DNA fragments were then extracted with H<sub>2</sub>O-saturated ether after having adjusted the volume to 500  $\mu$ l with TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). After extraction with 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1) and 1 volume of chloroform:isoamyl alcohol (24:1), the DNA fragments were precipitated in the presence of 1/10 volume of 3 M sodium acetate pH 6.0 and 2 volumes of absolute ethanol at -80°C for 30 min, centrifuged at 14000 g at 4°C for 30 min, washed with 70% ethanol, centrifuged at 14000 g at 4°C for 10 min, dried, taken up in 50  $\mu$ l of H<sub>2</sub>O. Next, 25  $\mu$ l of these DNA fragments were doubly digested with HindIII and EcoRI, purified by electrophoresis on 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the HindIII and EcoRI sites of the plasmid pNEB193 marketed by New England Biolabs, and previously dephosphorylated with the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100  $\mu$ g/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and )

Moved to page 28  
28

Moved to page 27 of the  
Substrate Specification

End page 27

( analyzed by enzymatic digestion. The resulting clone was called pBIOC44. The nucleotide sequence of the cDNA encoding the recombinant  $\alpha$  globin chain was verified by sequencing with the aid of the "Sequenase Version 2.0 DNA Sequencing" kit marketed by United States Biochemical (USB) as described in section I.b. The sequencing revealed two silent mutations situated at the forty-eighth nucleotide (C modified to T) and at the fifty-fourth (T modified to C) of the coding sequence for the  $\alpha$  globin chain.

*Moved to page 28 of the Substrate Specification*

b. CONSTRUCTION OF THE PLASMID pBIOC45 CONTAINING THE cDNA ENCODING  $\beta$  GLOBIN FOR CYTOPLASMIC TARGETING.

To obtain cytoplasmic targeting of the  $\beta$  globin chain, the methionine codon was fused with the first codon of the mature  $\beta$  globin chain by maintaining the open reading frame since ATG had been deleted from the construct M13mp10 cIIIFX beta.

The cDNA encoding the cytoplasmic targeting  $\beta$  globin chain was obtained by PCR amplification of the 146 codons constituting the mature  $\beta$  globin chain on the phage M13mp10 cIIIFX beta with the aid of the 2 oligodeoxynucleotides WD15 (5' gtcattaaatataaca ATG GTG CAC CTG ACT CCT GAG GAG AAG TCg GCC GTT AC 3') and WD16 (5' aatgagctcgtaacgcgt TTA GTG ATA CTT GTG GGC CAG GGC 3'). The WD15 primer provides the PacI restriction site, the aaca sequence favoring the initiation of translation (Joshi, 1987) and the initiator ATG codon followed by the first 12 codons of the mature  $\beta$  globin chain of which the ninth (TCT) is substituted for TCg (silent mutation) in order to create the EagI restriction site. The WD16 primer provides the MluI, HpaI and SacI restriction sites placed after the stop codon. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a. Next, 25  $\mu$ l of these DNA fragments were doubly digested with PacI and SacI, purified by electrophoresis on a 1.8% agarose gel )

Moved to page 29  
of the Subsection  
Specification

- 34 -

(and by the action of the "Geneclean II" kit (BIO101) and cloned at the PacI and SacI sites of the plasmid pNEB193 marketed by New England Biolabs, previously dephosphorylated by the enzyme calf intestinal alkaline phosphatase (New England) (Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100 µg/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC45. The nucleotide sequence of the cDNA encoding the recombinant  $\beta$  globin chain was verified by sequencing as described in section I.b.

C. CONSTRUCTION OF THE EXPRESSION BINARY PLASMIDS pBIOC46 AND pBIOC47, AND OF THE CO-EXPRESSION BINARY PLASMID pBIOC49 FOR CYTOPLASMIC TARGETING.

c.1. CONSTRUCTION OF THE BINARY PLASMID pBIOC46 CONTAINING cDNA ENCODING  $\alpha$  GLOBIN FOR CYTOPLASMIC TARGETING.

Starting with pBIOC44, the HindIII-EcoRI fragment carrying the cDNA encoding the cytoplasmic targeting  $\alpha$  globin chain was isolated by double enzymatic digestion with HindIII and EcoRI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101). Next, this DNA fragment was ligated with the plasmid DNA of pBIOC21 doubly digested with HindIII and EcoRI, and dephosphorylated with the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 10 µg/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The )

Moved to page 29  
of the Subsection  
Specification

resulting clone was called pBIOC46. The nucleotide sequence of the cDNA encoding the recombinant  $\alpha$  globin chain was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC46 was introduced by direct transformation into the Agrobacterium tumefaciens LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone selected was verified by enzymatic digestion of the plasmid DNA introduced.) end page 29

( c.2. CONSTRUCTION OF THE BINARY PLASMID pBIOC47 CONTAINING THE cDNA ENCODING  $\beta$  GLOBIN FOR CYTOPLASMIC TARGETING.

Starting with pBIOC45, the HindIII-EcoRI fragment carrying the cDNA encoding the cytoplasmic targeting  $\beta$  globin chain was isolated by double enzymatic digestion with HindIII (total digestion) and EcoRI (partial digestion), purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101). Next, this DNA fragment was ligated with the plasmid DNA of pBIOC21 doubly digested with HindIII and EcoRI and dephosphorylated with the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b.

The plasmid DNA of the clones obtained, selected on 10  $\mu$ g/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC47. The nucleotide sequence of the cDNA encoding the recombinant  $\beta$  globin chain was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC47 was introduced by direct transformation into the Agrobacterium tumefaciens LBA4404 strain according to the method of Holsters et al. (1978). The validity )

Moved to page 30  
of the substitute  
specification

- 36 -

of the clone selected was verified by enzymatic digestion of the plasmid DNA introduced.

c.3. CONSTRUCTION OF THE CO-EXPRESSION BINARY PLASMID pBIOC49 CONTAINING THE cDNA ENCODING THE  $\alpha$  AND  $\beta$  GLOBINS FOR CYTOPLASMIC TARGETING.

The HindIII-EcoRI fragment carrying the cDNA encoding the cytoplasmic targeting  $\alpha$  globin chain was isolated from pBIOC44 described in section II.c.1., and ligated with the plasmid DNA of pBIOC43 doubly digested with HindIII and EcoRI, and previously dephosphorylated with the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 12  $\mu\text{g}/\text{ml}$  tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC48.

The PacI-MluI fragment carrying the cDNA encoding the cytoplasmic targeting  $\beta$  globin chain was isolated from pBIOC45 described in section II.c.2., and ligated with the plasmid DNA of pBIOC48 doubly digested with PacI and MluI, and dephosphorylated by the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b., except that the E. coli Sure tet<sup>r</sup> strain was used in place of DH5 $\alpha$ . The Sure tet<sup>r</sup> strain is derived from the Sure strain (Stratagene) made sensitive to tetracycline by the loss of the F' episome. The plasmid DNA of the clones obtained, selected on 10  $\mu\text{g}/\text{ml}$  tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC49. )

Moved to page 31  
of the substitute specification

end page 30

Moved to page 31 of  
The Substitution Specification

- 37 -

( The nucleotide sequence of the cDNAs encoding the recombinant  $\alpha$  and  $\beta$  globin chains was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC49 was introduced by direct transformation into the Agrobacterium tumefaciens LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone selected was verified by enzymatic digestion of the plasmid DNA introduced.

### III. CONSTRUCTION OF THE CHIMERIC GENES ENCODING THE $\alpha$ AND $\beta$ GLOBIN CHAINS ALLOWING EXPRESSION OF RECOMBINANT HUMAN HEMOGLOBIN IN THE MITOCHONDRIA OF TOBACCO LEAVES

To obtain mitochondrial targeting, the sequence encoding the transit peptide of the Nicotiana plumbaginifolia mitochondrial ATPase-F1  $\beta$  subunit precursor (ATG GCT TCT CGG AGG CTT CTC GCC TCT CTC CTC CGT CAA TCG GCT CAA CGT GGC GGC GGT CTA ATT TCC CGA TCG TTA GGA AAC TCC ATC CCT AAA TCC GCT TCA CGC GCC TCT TCA CGC GCA TCC CCT AAG GGA) TTC CTC TTA AAC CGC GCC GTA CAG TAC) is fused with the first codon of the sequence encoding, on the one hand, the mature  $\alpha$  globin chain (deletion of the initiator ATG) and, on the other hand, the mature  $\beta$  globin chain (deletion of the initiator ATG) while maintaining the open reading frames.

The sequence encoding the Nicotiana plumbaginifolia mitochondrial ATPase F1  $\beta$  subunit is contained in the plasmid pTZ-catp2-1 provided by Boutry. This plasmid corresponds to the plasmid pTZ18R containing the cDNA (cNP10) as described by Boutry and Chua (1985).

The N-terminal transit peptide, composed of 54 amino acids as defined by Chaumont et al. (1994), was used during the carrying out of the constructions.)

Moved to Page 32  
of the Substitution  
Specification

( a. CONSTRUCTION OF THE PLASMID pBIOC50 CONTAINING THE  
DNA ENCODING  $\alpha$  GLOBIN FOR MITOCHONDRIAL TARGETING.

To obtain mitochondrial targeting of the  $\alpha$  globin chain, the sequence encoding the transit peptide of the Nicotiana plumbaginifolia mitochondrial ATPase-F1  $\beta$  subunit precursor was fused with the first codon of the sequence encoding the mature  $\alpha$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the transit peptide and the mature  $\alpha$  globin chain is Tyr-Val.

The sequence encoding the transit peptide of the mitochondrial ATPase-F1  $\beta$  subunit precursor was amplified by PCR on the plasmid pTZ-catp2-1 with the aid of the 2 oligodeoxynucleotides, WD17 (5' cgcaagcttaaca ATG GCT TCT CGG AGG CTT CTC 3') and WD18 (5' tag aat tc GGC cGG AGA CAG CAC GTA CTG TAC GGC GCG GTT TAA G 3'). The WD17 primer provides the HindIII restriction site, the aaca sequence promoting the initiation of translation (Joshi, 1987) and the first 7 codons of the transit peptide (including the initiator ATG). The WD18 primer provides the EcoRI restriction site, the first 5 codons of the sequence encoding the mature  $\alpha$  globin chain (an EagI restriction site is created by silent mutation in the fourth codon (CCT modified to CCg) and the last 7 codons of the sequence of the transit peptide. PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third step of chapter II.a. Next, these DNA fragments were doubly ) ( digested with HindIII and EagI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the HindIII and EagI sites of the plasmid pBIOC44 described in section II.a., previously purified by electrophoresis on a 0.75% agarose gel and using the "Geneclean II" kit. The plasmid pBIOC44 was dephosphorylated by the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as )

Moved to page 32  
of the substack  
Spec'f actions

Moved to page  
33 of substack  
Spec'f actions

(described in section I. b. The plasmid DNA of the clones obtained, selected on 100 µg/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC50. The nucleotide sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the transit peptide and the cDNA encoding the mature  $\alpha$  globin chain was verified by sequencing as described in section I.b. The sequencing revealed two silent mutations situated at the tenth nucleotide (C modified to A) and at the one hundred forty first (C modified to G) of the coding sequence for the transit peptide.

b. CONSTRUCTION OF THE PLASMID pBIOC51 CONTAINING THE cDNA ENCODING  $\beta$  GLOBIN FOR MITOCHONDRIAL TARGETING.

To obtain mitochondrial targeting of the  $\beta$  globin chain, the sequence encoding the transit peptide of the Nicotiana plumbaginifolia mitochondrial ATPase-F1  $\beta$  subunit precursor was fused with the first codon of the sequence encoding the mature  $\beta$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the transit peptide and the mature  $\beta$  globin chain is Tyr-Val.

The sequence encoding the transit peptide of the mitochondrial ATPase-F1  $\beta$  subunit precursor was amplified by PCR on the plasmid pTZ-catp2-1 with the aid of the 2 oligodeoxynucleotides, WD19 (5' gtcattattaaca ATG GCT TCT CGG AGG CTT CTC GCC TCT C 3') and WD20 (5' aatgagct C GGC cGA CTT CTC CTC AGG AGT CAG GTG CAC GTA CTG TAC GGC GCG GTT TAA G 3'). The WD19 primer provides the PacI restriction site, the aaca sequence promoting the initiation of translation (Joshi, 1987) and preceding the first 9 codons of the transit peptide (including the initiator ATG). The WD20 primer provides the SacI restriction site, the first 10 codons of the sequence encoding the mature  $\beta$  globin chain (an EagI restriction site is created by silent)

moved to page 34 of the submitted specification

-40-

(mutation in the nineth codon (TCT modified to TCg)) and the last 7 codons of the sequence of the transit peptide. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a. Next, these DNA fragments were doubly digested with PacI and EagI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the PacI and EagI sites of the plasmid pBIOC45 described in section II.b., previously purified by electrophoresis on a 0.75% agarose gel and using the "Geneclean II" kit. The plasmid pBIOC45 was dephosphorylated by the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100 µg/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC51. The nucleotide sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the transit peptide and the cDNA encoding the mature  $\beta$  globin chain was verified by sequencing as described in section I.b.

c. CONSTRUCTION OF THE CO-EXPRESSION BINARY PLASMID pBIOC53 CONTAINING THE cDNAs ENCODING THE  $\alpha$  AND  $\beta$  GLOBINS, FOR MITOCHONDRIAL TARGETING.

The HindIII-EcoRI fragment carrying the cDNA encoding the mitochondrial targeting  $\alpha$  globin chain was isolated from pBIOC50 described in section III.a., and ligated to the plasmid DNA of pBIOC43 doubly digested with HindIII and EcoRI, and dephosphorylated by the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on )

moved to page  
34 of the  
Substrate  
Specification

-41-

10 µg/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC52.) end page 34

(The PacI-MluI fragment carrying the cDNA encoding the mitochondrial targeting  $\beta$  globin chain was isolated from pBIOC51 described in section III.b., and ligated to the plasmid DNA of pBIOC52 doubly digested with PacI and MluI, and dephosphorylated by the enzyme calf intestinal alkaline phosphatase (New England Biolabs). The ligation and the transformation were carried out as described in section II.c.3 using the E. coli Sure tet<sup>r</sup> strain. The plasmid DNA of the clones obtained, selected on 10 µg/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC53.

The nucleotide sequence of the cDNAs encoding the recombinant  $\alpha$  and  $\beta$  globin chains allowing a mitochondrial targeting was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC53 was introduced by direct transformation into the Agrobacterium tumefaciens LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone obtained was verified by enzymatic digestion of the plasmid DNA introduced.

**IV. CONSTRUCTION OF THE CHIMERIC GENES ENCODING THE  $\alpha$  AND  $\beta$  GLOBIN CHAINS ALLOWING EXPRESSION OF RECOMBINANT HUMAN HEMOGLOBIN IN THE CHLOROPLASTS OF TOBACCO LEAVES.**

To obtain chloroplastic targeting, the sequence encoding the transit peptide of the precursor of the small subunit of ribulose 1,5-diphosphate carboxylase of Pisum sativum L. (ATG GCT TCT ATG ATA TCC TCT TCA GCT GTG ACT ACA GTC AGC CGT GCT TCT ACG GTG CAA TCG GCC GCG GTG GCT CCA TTC GGC GGC CTC AAA TCC ATG ACT GGA TTC CCA GTT AAG AAG GTC AAC ACT GAC ATT ACT TCC ATT ACA AGC )

moved to page 35 of the Substrate Specification

Moved to  
page 35 of  
the submitted  
specification

-42-

AAT GGT GGA AGA GTA AAG TGC) is fused with the first codon of the sequence encoding, on the one hand, the mature  $\alpha$  globin chain (deletion of the initiator ATG) and, on the other hand, the mature  $\beta$  globin chain (deletion of the initiator ATG) while maintaining the open reading frames.) end page 35

(This N-terminal transit peptide, composed of 57 amino acids, as defined by Anderson et al. (1986), was isolated from the plasmid pJIT117 (Guerineau et al., 1988) and used during the carrying out of the constructions.

a. CONSTRUCTION OF THE PLASMID pBIOC55 CONTAINING THE cDNA ENCODING  $\alpha$  GLOBIN FOR CHLOROPLAST TARGETING.

To obtain chloroplast targeting of the  $\alpha$  globin chain, the sequence encoding the transit peptide of the precursor of the small subunit of the ribulose 1,5-diphosphate carboxylase of Pisum sativum L. was fused with the first codon of the sequence encoding the mature  $\alpha$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the transit peptide and of the mature  $\alpha$  globin chain is Cys-Val.

The sequence of the transit peptide of the precursor of the small subunit of ribulose 1,5-diphosphate carboxylase was amplified by PCR on the plasmid pJIT117 with the aid of the 2 oligodeoxynucleotides, WD21 (5' cgcaagcttaaca ATG GCT TCT ATG ATA TCC TCT TCA GC 3') and WD22 (5' tag aat tC GGC cGG AGA CAG CAC GCA CTT TAC TCT TCC ACC ATT GC 3'). The WD21 primer provides the HindIII restriction site, the aaca sequence promoting the initiation of translation (Joshi, 1987) and the first 8 codons of the transit peptide (including the initiator ATG). The WD22 primer provides the EcoRI restriction site, the first 5 codons of the sequence encoding the mature  $\alpha$  globin chain (an EagI restriction site is created by silent mutation in the fourth codon (CCT modified to CCg)) and the last 7 codons of the sequence of the transit )

peptide. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a.

Next, these DNA fragments were doubly digested with HindIII and EcoRI and cloned at the HindIII and EcoRI sites of the plasmid pNEB193 marketed by New England Biolabs. The plasmid pNEB193 was dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100 µg/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC54. The nucleotide sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the transit peptide and the cDNA encoding the mature  $\alpha$  globin chain was verified by sequencing as described in section I.b.)

From the plasmid pBIOC54, the HindIII-EagI fragment, carrying the sequence encoding the transit peptide of the precursor of the small subunit of ribulose 1,5-diphosphate carboxylase and the first 4 codons of the mature  $\alpha$  globin chain was isolated by double digestion, HindIII (total digestion) and EagI (partial digestion). This HindIII-EagI fragment, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) was cloned at the HindIII and EagI sites of the dephosphorylated plasmid pBIOC44 as described in section II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100 µg/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC55. The nucleotide sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the transit peptide and the cDNA >

Moved to

Page 36

of the subset of the specification

Moved to Page 37 of the  
subset of the specification

Moved to page 37 of  
the substitute specification

-44-

encoding the mature  $\alpha$  globin chain was verified by sequencing as described in section I.b.

b. CONSTRUCTION OF THE PLASMID pBIOC57 CONTAINING THE cDNA ENCODING  $\beta$  GLOBIN FOR CHLOROPLAST TARGETING.

To obtain chloroplast targeting of the  $\beta$  globin chain, the sequence encoding the transit peptide of the Pisum sativum L. ribulose 1,5-diphosphate carboxylase small subunit precursor was fused with the first codon of the sequence encoding the mature  $\beta$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the transit peptide and the mature  $\beta$  globin chain is Cys - Val.

The sequence encoding the transit peptide of the ribulose 1,5-diphosphate carboxylase small subunit precursor was amplified by PCR on the plasmid pJIT117 with the aid of the 2 oligodeoxynucleotides, WD23 (5' gtcattaaatcca ATG GCT TCT ATG ATA TCC TCT TCA GCT GTG 3') and WD24 (5' aatgagct C GGC cGA CTT) (CTC CTC AGG AGT CAG GTG CAC GCA CTT TAC TCT TCC ACC 3'). The WD23 primer provides the PacI restriction site, the aaca sequence promoting the initiation of translation (Joshi, 1987) and preceding the first 10 codons of the transit peptide (including the initiator ATG). The WD24 primer provides the SacI restriction site, the first 10 codons of the sequence encoding the mature  $\beta$  globin chain (an EagI restriction site is created by silent mutation in the ninth codon (TCT modified to TCg)) and the last 6 codons of the sequence of the transit peptide. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a. Next, these DNA fragments were doubly digested with PacI and SacI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the PacI and SacI sites of the plasmid pNEB193 marketed by New England Biolabs. The plasmid pNEB193 was dephosphorylated as described in II.a. The )

Moved to page 38 of the  
substitute specification

ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100 µg/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC56. The nucleotide sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the transit peptide and the cDNA encoding the mature  $\beta$  globin chain was verified by sequencing as described in section I.b.

From the plasmid pBIOC56, the PacI-EagI fragment, carrying the sequence of the transit peptide of the ribulose 1,5-diphosphate carboxylase small subunit precursor and the first 9 codons of the sequence encoding the mature  $\beta$  globin chain, was isolated by double digestion, PacI (total digestion) and EagI (partial digestion). This PacI-EagI fragment, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101), was cloned at the PacI and EagI sites of the dephosphorylated plasmid pBIOC45 as described in section II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100 µg/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990), and analyzed by enzymatic digestion. The resulting clone was called pBIOC57. The nucleotide sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the transit peptide and the cDNA encoding the mature  $\beta$  globin chain was verified by sequencing as described in section I.b.)

→ moved to  
page 39 of  
Substitute  
specification

Moved to page 38  
of the substitute  
specification

( c. CONSTRUCTION OF THE CO-EXPRESSION BINARY PLASMID  
pBIOC59 CONTAINING THE cDNAs ENCODING THE  $\alpha$  AND  $\beta$   
GLOBINS, FOR CHLOROPLAST TARGETING.

The HindIII-EcoRI fragment carrying the cDNA encoding the chloroplast targeting  $\alpha$  globin chain was isolated from pBIOC55 described in section IV.a., and ligated to the plasmid DNA of pBIOC43 doubly digested with HindIII and EcoRI and dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 10  $\mu$ g/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC58.

The PacI-MluI fragment carrying the cDNA encoding the chloroplast targeting  $\beta$  globin chain was isolated from pBIOC57 described in section IV.b., and ligated to the plasmid DNA of pBIOC58 doubly digested with PacI and MluI, and dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section II.c.3 using the *E. coli* Sure tet<sup>r</sup> strain. The plasmid DNA of the clones obtained, selected on 10  $\mu$ g/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC59.

The nucleotide sequence of the cDNAs encoding the recombinant  $\alpha$  and  $\beta$  globin chains allowing chloroplast targeting was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC59 was introduced by direct transformation into the *Agrobacterium tumefaciens* LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone selected was verified by enzymatic digestion of the plasmid DNA introduced. ) end page 39

Moved to page 39 of specification  
of the subspecies

- 17 -

v. CONSTRUCTION OF THE CHIMERIC GENES ENCODING THE  $\alpha$  AND  $\beta$  GLOBIN CHAINS ALLOWING EXPRESSION OF THE RECOMBINANT HUMAN HEMOGLOBIN FOR SECRETION IN TOBACCO LEAVES.

To obtain secretion, the sequence encoding the signal peptide (SP) of sporamine A of the tuberized roots of sweet potato (Murakami et al., 1986; Matsuoka and Nakamura, 1991) (ATG AAA GCC TTC ACA CTC GCT CTC TTC TTA GCT CTT TCC CTC TAT CTC CTG CCC AAT CCA GCC CAT TCC), is fused with the first codon of the sequence encoding, on the one hand, the mature  $\alpha$  globin chain (deletion of the initiator ATG) and, on the other hand, the mature  $\beta$  globin chain (deletion of the initiator ATG) while maintaining the open reading frames. This signal peptide of 23 amino acids was isolated from the plasmid pMAT103 (Matsuoka and Nakamura, 1991) and used during the carrying out of the constructions.

a. CONSTRUCTION OF THE PLASMID pBIOC60 CONTAINING THE cDNA ENCODING  $\alpha$  GLOBIN FOR SECRETION...

To obtain the secretion of the  $\alpha$  globin chain, the sequence encoding the signal peptide of sweet potato sporamine A was fused with the first codon of the mature  $\alpha$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the signal peptide and the mature  $\alpha$  globin chain is Ser-Val.

The sequence encoding the signal peptide (SP) of the sporamine A of the tuberized roots of sweet potato was amplified by PCR on the plasmid pMAT103 with the aid of 2 oligodeoxynucleotides, WD25 (5' cgcaagcttaaca ATG AAA GCC TTC ACA CTC GC 3') and WD26 (5' tagaattC GGC cGG AGA CAG CAC GGA ATG GGC TGG ATT GGG CAG G 3'). The WD25 primer provides the HindIII restriction site, the aaca sequence promoting the initiation of translation (Joshi, 1987) and the first 6 codons of the signal peptide (including the initiator )

40 of the secreted protein

Moved to page

Moved to page  
40 of my  
scattered  
specimens

ATG). The WD26 primer provides the EcoRI restriction site, the first 5 codons of the sequence encoding the mature  $\alpha$  globin chain (an EagI restriction site is created by silent mutation in the fourth codon) ((CCT modified to CCg)) and the last 7 codons of the sequence of the signal peptide. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a. Next, these DNA fragments were doubly digested with HindIII and EagI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the HindIII and EagI sites of the dephosphorylated plasmid pBIOC44 described in section II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100  $\mu$ g/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC60. The nucleotide sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the signal peptide and the cDNA encoding the mature  $\alpha$  globin chain was verified by sequencing as described in section I.b.

b. CONSTRUCTION OF THE PLASMID pBIOC61 CONTAINING THE cDNA ENCODING THE  $\beta$  GLOBIN FOR SECRETION.

To obtain the secretion of the  $\beta$  globin chain, the sequence encoding the signal peptide of the sweet potato sporamine A was fused with the first codon of the mature  $\beta$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the signal peptide and the mature  $\beta$  globin chain is Ser-Val.

The sequence encoding the signal peptide (SP) of the sporamine A of the tuberized roots of sweet potato was amplified by PCR on the plasmid pMAT103 with the aid of the 2 oligodeoxynucleotides, WD27 )

Moved to page 41 of  
the Sweet Potato  
Specimens

(5' gtcattaaatcca ATG AAA GCC TTC ACA CTC GC 3') and WD28 (5' aatgagct C GGC cGA CTT CTC CTC AGG AGT CAG GTG CAC GGA ATG GGC TGG ATT GGG CAG G 3'). The WD27 primer provides the PacI restriction site, the aaca sequence promoting the initiation of translation (Joshi, 1987) and the first 6 codons of the signal peptide (including the initiator ATG). The WD28 primer provides the SacI restriction site, the first 10 codons of the sequence encoding the mature  $\beta$  globin chain (an EagI site is created by silent mutation in the ninth codon (TCT modified to TCg)) and the last 7 codons of the sequence of the signal peptide. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a. Next, these DNA fragments were doubly digested with PacI and EagI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the PacI and EagI sites of the dephosphorylated plasmid pBIOC45 described in section II.b. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100  $\mu$ g/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC61. The nucleotide sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the signal peptide and the cDNA encoding the mature  $\beta$  globin chain was verified by sequencing as described in section I.b.

C. CONSTRUCTION OF THE CO-EXPRESSION BINARY PLASMID pBIOC63 CONTAINING THE cDNAs ENCODING THE  $\alpha$  AND  $\beta$  GLOBINS FOR SECRETION

The HindIII-EcoRI fragment carrying the cDNA encoding the  $\alpha$  globin chain for secretion was isolated from pBIOC60 described in section V.a., and ligated to the plasmid DNA of pBIOC43 doubly digested with HindIII )

Moved to  
Page 41

Specification  
of the substituent  
to page 42

Moved to

Moved to page 42  
in the subsection Specification

-50-

(and EcoRI, and dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section II.c.3 using the E. coli Sure tet. strain. The plasmid DNA of the clones obtained, selected on 10 µg/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC62.

The PacI-MluI fragment carrying the cDNA encoding the β globin chain for secretion was isolated from pBIOC61 described in section V.b., and ligated to the plasmid DNA of pBIOC62 doubly digested with PacI and MluI, and dephosphorylated as described in II.a. The ligation and the transformation were carried out as described) (in section I.b. The plasmid DNA of the clones obtained, selected on 10 µg/ml tetracyclin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC63.

The nucleotide sequence of the cDNAs encoding the α and β globin chains for secretion was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC63 was introduced by direct transformation into the Agrobacterium tumefaciens LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone selected was verified by enzymatic digestion of the plasmid DNA introduced.

## VI. CONSTRUCTION OF THE CHIMERIC GENES ENCODING THE α AND β GLOBIN CHAINS ALLOWING EXPRESSION OF RECOMBINANT HUMAN HEMOGLOBIN IN THE ENDOPLASMIC RETICULUM OF TOBACCO LEAVES.

The sequence encoding the KDEL signal (Lys-Asp-Glu-Leu), placed at the C-terminal end of the α and β globin chains upstream of the stop codon combined with the presence of the sequence encoding the N-terminal signal peptide (SP) of sporamine A of the tuberized

Moved to page 43 in the subsection Specification

roots of sweet potato allows targeting in the endoplasmic reticulum.

Moved to page 43  
in the subsection  
Specification

a. CONSTRUCTION OF THE PLASMID pBIOC65 CONTAINING THE cDNA ENCODING THE  $\alpha$  GLOBIN ALLOWING RETENTION IN THE ENDOPLASMIC RETICULUM.

To obtain retention in the endoplasmic reticulum, the sequence encoding the KDEL signal (5' aaa gat gag cta 3') was placed before the first stop codon (TGA) of the mature  $\alpha$  globin chain while maintaining the open reading frame.

The plasmid containing the cDNA encoding the  $\alpha$  globin chain which contains the sequence encoding the KDEL signal placed before its first stop codon was obtained by following the same steps as for the manufacture of the plasmid pBIOC44 (described in II.a. except that the WD29 primer (5' gcgaattc TCA tag ctc atc ttt ACG GTA TTT GGA GGT CAG CAC 3') replaces the WD14 primer. The WD29 primer provides the EcoRI restriction site and the KDEL sequence situated respectively after and before the stop codon.

The resulting plasmid obtained was called pBIOC64. The nucleotide sequence of the chimeric gene between the cDNA encoding the  $\alpha$  globin chain and the sequence encoding the  $\alpha$  KDEL signal was verified by sequencing as described in section I.b.

Next, the plasmid pBIOC64 was modified as described in V.a. by translational fusion with the signal peptide of sporamine A of the tuberized roots of sweet potato to give the plasmid pBIOC65 allowing targeting in the endoplasmic reticulum. The nucleotide sequence of the chimeric gene between the sequence encoding the signal peptide, the cDNA encoding the mature  $\alpha$  globin chain and the sequence encoding the KDEL signal was verified by sequencing as described in section I.b. The cleaving sequence between the sequences of the signal peptide and the mature  $\alpha$  globin chain is Ser - Val. )

Moved to page 44 of the  
Substitution Specification

moved to page 44  
of the substituted  
specification

-52-

b. CONSTRUCTION OF THE PLASMID pBIOC67 CONTAINING THE cDNA ENCODING THE  $\beta$  GLOBIN ALLOWING RETENTION IN THE ENDOPLASMIC RETICULUM.

To obtain retention in the endoplasmic reticulum, the sequence encoding the KDEL signal (5' aaa gat gag cta 3') was placed before the first stop codon (TAA) of the mature  $\beta$  globin chain while maintaining the open reading frame.

The plasmid containing the cDNA encoding the  $\beta$  globin chain which contains the sequence encoding the KDEL signal before its first stop codon was obtained by following the same steps as for the manufacture of the plasmid pBIOC45 described in II.b., except that the WD30 primer (5' aatgagctcgtaacgcgt TTA tag ctc atc ttt GTG ATA CTT GTG GGC CAG GGC 3') replaces the WD16 primer. The WD30 primer provides the MluI, HpaI and SacI restriction sites and the KDEL sequence placed respectively after and before the stop codon.) *end part 44*

(The resulting plasmid obtained was called pBIOC66. The nucleotide sequence of the chimeric gene between the cDNA encoding the  $\beta$  globin chain and the sequence encoding the KDEL signal was verified by sequencing as described in section I.b.

Next, the plasmid pBIOC66 was modified as described in V.b. by translational fusion with the signal peptide of the sporamine A of the tuberized roots of sweet potato to give the plasmid pBIOC67 allowing targeting in the endoplasmic reticulum. The nucleotide sequence of the chimeric gene between the sequence encoding the signal peptide, the cDNA encoding the mature  $\beta$  globin chain and the sequence encoding the KDEL signal was verified by sequencing as described in section I.b. The cleaving sequence between the sequences of the signal peptide and the mature  $\beta$  globin chain is Ser - Val. )

moved to page 45  
of the substituted  
specification

Moved to page 45  
of the subsection  
specification

-53-

C. CONSTRUCTION OF THE CO-EXPRESSION BINARY PLASMID pBIOC69 CONTAINING THE cDNAs ENCODING THE  $\alpha$  AND  $\beta$  GLOBINS ALLOWING RETENTION IN THE ENDOPLASMIC RETICULUM.

The HindIII-EcoRI fragment carrying the cDNA encoding the  $\alpha$  globin chain allowing retention in the endoplasmic reticulum was isolated from pBIOC65 described in section VI.a., and ligated to the plasmid DNA of pBIOC43 doubly digested with HindIII and EcoRI, and dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 10  $\mu$ g/ml tetracyclin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC68.

The PacI-MluI fragment carrying the cDNA encoding the  $\beta$  globin chain allowing retention in the endoplasmic reticulum was isolated from pBIOC67 described in section VI.b., and ligated to the plasmid DNA of pBIOC68 doubly digested with PacI and MluI, and dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section II.c.3 using the E. coli Sure tet<sup>r</sup> strain. The plasmid DNA of the clones obtained, selected on 10  $\mu$ g/ml tetracyclin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC69.

The nucleotide sequence of the cDNAs encoding the  $\alpha$  and  $\beta$  globin chains allowing their retention in the endoplasmic reticulum was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC69 was introduced by direct transformation into the Agrobacterium tumefaciens LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone selected was verified by enzymatic digestion of the plasmid DNA introduced.)

Moved to page 46  
of the subsection  
specification

moved to page 46  
of the Substituted Specification

-54-

## VII. CONSTRUCTION OF THE CHIMERIC GENES ENCODING THE $\alpha$ AND $\beta$ GLOBIN CHAINS ALLOWING EXPRESSION OF THE RECOMBINANT HUMAN HEMOGLOBIN IN THE VACUOLES OF TOBACCO LEAVES.

To allow vacuolar targeting, the sequence encoding the prepropeptide (PPS) of sporamine A of the tuberized roots of sweet potato (Murakami et al., 1986; Matsuoka and Nakamura, 1991), which corresponds to the signal peptide followed by the N-terminal sequence for vacuolar targeting (ATG AAA GCC TTC ACA CTC GCT CTC TTC TTA GCT CTT TCC CTC TAT CTC CTG CCC AAT CCA GCC CAT TCC AGG TTC AAT CCC ATC CGC CTC CCC ACC ACA CAC GAA CCC GCC), is fused with the first codon of the sequence encoding, on the one hand, the mature  $\alpha$  globin chain (deletion of the initiator ATG) and, on the other hand, the mature  $\beta$  globin chain (deletion of the initiator ATG) while maintaining the open reading frames. This prepropeptide of 37 amino acids was isolated from the plasmid pMAT103 (Matsuoka and Nakamura, 1991) and used during the carrying out of the constructions.

To obtain vacuolar targeting of the  $\alpha$  globin chain, the sequence encoding the prepropeptide of sweet potato sporamine A was fused with the first codon of the mature  $\alpha$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the signal peptide and the mature  $\alpha$  globin chain is Ala-Val.) end page 46

### a. CONSTRUCTION OF THE PLASMID pBILOC70 CONTAINING THE cDNA ENCODING THE $\alpha$ GLOBIN ALLOWING VACUOLAR TARGETING.

moved to  
page 47 of  
Substituted Specification

The sequence encoding the N-terminal prepropeptide (PPS) of the sporamine A of the tuberized roots of sweet potato was amplified by PCR on the plasmid pMAT103 with the aid of the 2 oligodeoxy-nucleotides, WD25 (5' cgcaagcttaaca ATG AAA GCC TTC ACA CTC GC 3') described in v.a. and WD31 (5' tagaattc GGC CGG AGA CAG CAC GGC GGG TTC GTG TGT GGT TG 3'). The

MOVED TO PAGE 47  
OF THE SUBSTITUTED SPECIFICATION

- 13 -

WD31 primer provides the EcoRI restriction site, the first 5 codons of the sequence encoding the mature  $\alpha$  globin chain (an EagI site is created by silent mutation in the fourth codon (CCT modified to CCg)) and the last 6 codons of the sequence of the N-terminal prepropeptide. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a. Next, these DNA fragments were doubly digested with HindIII and EagI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the HindIII and EagI sites of the dephosphorylated plasmid pBIOD44 described in section II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100  $\mu$ g/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOD70. The nucleotide sequence of the chimeric gene between the sequence encoding the prepropeptide and the cDNA encoding the mature  $\alpha$  globin chain was verified by sequencing as described in section I.b.

b. CONSTRUCTION OF THE PLASMID pBIOD71 CONTAINING THE cDNA ENCODING  $\beta$  GLOBIN ALLOWING VACUOLAR TARGETING.

To obtain vacuolar targeting of the  $\beta$  globin chain, the sequence encoding the prepropeptide of sporamine A of sweet potato was fused with the first codon of the mature  $\beta$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the signal peptide and the mature  $\beta$  globin chain is Ala-Val.

The sequence encoding the N-terminal prepropeptide (PPS) of sporamine A of the tuberized roots of sweet potato was amplified by PCR on the plasmid pMAT103 with the aid of the 2 oligodeoxy-nucleotides, WD27 (5' gtcattaaatataaaca ATG AAA GCC TTC

) MOVED TO PAGE 48  
OF THE SUBSTITUTED SPECIFICATION

-50-

(ACA CTC GC 3') described in V.b. and WD32 (5' aatgagcc  
C GGC cGA CTT CTC CTC AGG AGT CAG GTG CAC GGC GGG TTC  
GTG TGT GGT TG 3'). The WD32 primer provides the SacI  
restriction site, the first 10 codons of the sequence  
encoding the mature  $\beta$  globin chain (an EagI restriction  
site is created by silent mutation in the ninth codon  
(TCT modified to TCg)) and the last 6 codons of the  
sequence of the N-terminal prepropeptide. The PCR  
amplification and the treatment of the amplified DNA  
fragments were carried out as described in the third  
stage of section II.a. Next, these DNA fragments were  
doubly digested with PacI and EagI, purified by  
electrophoresis on a 1.8% agarose gel and by the action  
of the "Geneclean II" kit (BIO101) and cloned at the  
PacI and EagI sites of the dephosphorylated plasmid  
pBIOC45 described in section II.b. The ligation and the  
transformation were carried out as described in section  
I.b. The plasmid DNA of the clones obtained, selected  
on 100  $\mu$ g/ml ampicillin, was extracted according to the  
alkaline lysis method (Stephen et al., 1990) and  
analyzed by enzymatic digestion. The resulting clone  
was called pBIOC71. The nucleotide sequence of the  
chimeric gene between the sequence encoding the  
prepropeptide and the cDNA encoding the mature  $\beta$  globin  
chain was verified by sequencing as described in  
section I.b.

c. CONSTRUCTION OF THE CO-EXPRESSION BINARY PLASMID  
pBIOC73 CONTAINING THE cDNAs ENCODING THE  $\alpha$  AND  $\beta$   
GLOBINS ALLOWING VACUOLAR TARGETING.

The HindIII-EcoRI fragment carrying the cDNA  
encoding the vacuolar targeting  $\alpha$  globin chain was  
isolated from pBIOC70 described in section VII.a., and  
ligated to the plasmid DNA of pBIOC43 doubly digested  
with HindIII and EcoRI, and dephosphorylated as  
described in II.a. The ligation and the transformation  
were carried) out as described in section I.b. The  
plasmid DNA of the clones obtained, selected on)

Page 49  
of Substitution  
Specification

moved to page 48  
of the Substitution  
Specification

( 10 µg/ml tetracyclin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC72.

The PacI-MluI fragment carrying the cDNA encoding the vacuolar targeting  $\beta$  globin chain was isolated from pBIOC71 described in section VII.b., and ligated to the plasmid DNA of pBIOC72 doubly digested with PacI and MluI, and dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section II.c.3. using the *E. coli* Sure tet<sup>r</sup> strain. The plasmid DNA of the clones obtained, selected on 10 µg/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC73.

The nucleotide sequence of the cDNAs encoding the vacuolar targeting  $\alpha$  and  $\beta$  globin chains was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC73 was introduced by direct transformation into the *Agrobacterium tumefaciens* LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone selected was verified by enzymatic digestion of the plasmid DNA introduced.

## VIII: PRODUCTION OF TRANSGENIC TOBACCO PLANTS

The tobacco plants used for the transformation experiments (*Nicotiana tabacum* var. PBD6) are cultured in vitro on Murashige and Skoog basic medium (1962) supplemented with Gamborg et al. vitamins (1968, Sigma reference M0404), sucrose at 20 g/l and agar (Merck) at 8 g/l. The pH of the medium is adjusted to 5.8 with a solution of potassium hydroxide before autoclaving at 120°C for 20 min. The tobacco plantlets are transplanted by taking internode cuttings every 30 days on this MS20 propagation medium.)

Mixed w/ Page 49 of the subsequent  
selections

*Mixed stage 49*

-3-

( All the in vitro cultures are carried out in an air-conditioned chamber under the conditions defined below:

- Light intensity of  $30 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ ; photoperiod of 16 h;
- Thermoperiod of  $26^\circ\text{C}$  by day,  $24^\circ\text{C}$  by night.)  
*en page 49*

( The transformation technique used is derived from that of Horsch et al. (1985).

A preculture of *Agrobacterium tumefaciens* LBA4404 strain containing the plasmids pBIOC46 or pBIOC47 or pBIOC49 or pBIOC53 or pBIOC59 is carried out for 48 h at  $28^\circ\text{C}$ , with stirring, in LB medium supplemented with appropriate antibiotics (rifampicin and tetracycline). The preculture is then diluted 50-fold in the same medium and cultured under the same conditions. After one night, the culture is centrifuged (10 min, 3000 g), the bacteria are taken up in an equivalent volume of liquid MS30 medium (30 g/l sucrose) and this suspension is diluted 10-fold.

Explants of about  $1 \text{ cm}^2$  are cut from the leaves of the plantlets described above. They are then brought into contact with the bacterial suspension for 1 h, and then dried rapidly on filter paper and placed on a coculture medium (solid MS30).

After 2 days, the explants are transferred to Petri dishes on MS30 regeneration medium containing a selective agent, kanamycin (200 mg/l), a bacteriostatic, augmentin (400 mg/l) and the hormones necessary for the induction of buds (BAP, 1 mg/l and NAA, 0.1 mg/l). A transplantation of the explants is carried out on the same medium after 2 weeks of culture. After a further 2 weeks, the buds are transplanted into Petri dishes on the development medium composed of the MS20 medium supplemented with kanamycin and augmentin. After 15 days, the buds are transplanted into pots on the same medium whose kanamycin concentration has been decreased by one half. The rooting takes about 20 days, at the end of which )

*Mixed stage 50  
of the  
subculture  
specification*

Moved to page 50  
of the substitute  
specification

(the plantlets can be cloned using internode cuttings in vitro or taken out to the greenhouse.

#### IX: PARTIAL EXTRACTION AND PARTIAL PURIFICATION OF RECOMBINANT PROTEINS FROM TOBACCO LEAVES

Fifty grams of transformed tobacco leaves (fresh weight) are ground in liquid nitrogen and then left stirring for 15 min at 4°C in 300 ml of 50 mM tris-HCl buffer pH 8 supplemented with 1 mM EDTA, 1 mM β-mercaptoethanol and polyvinylpyrrolidone (PVP, 10 g/300 ml). The ground product is filtered on miracloth and then centrifuged for 20 min at 4°C at 10000 g. The supernatant is again filtered on miracloth. The proteins are then precipitated for 12 h at 4°C with a solution of ammonium sulfate at saturation. After centrifuging for 20 min at 10000 g, the pellet is taken up in 50 mM tris-HCl buffer pH 8 supplemented with 1 mM DTT and 1 mM EDTA and dialyzed twice 12 hours against this same buffer. After dialysis, the retentate is centrifuged and then filtered on miracloth. An assay of proteins is also carried out according to the Bradford technique (1976).

First purification step: Equilibration in 10 mM phosphate buffer pH 6.7-1 mM EDTA by passing over a Sephadex G25 resin and then loading onto an ion-exchange resin (CM cellulose) equilibrated in 10 mM phosphate buffer pH 6.7, 1 mM EDTA. Washing with 4 volumes of this same buffer and then eluting with a linear gradient from 10 mM Na<sub>2</sub>HPO<sub>4</sub> pH 6.7, 1 mM EDTA to 100 mM Na<sub>2</sub>HPO<sub>4</sub> pH 6.7, 1 mM EDTA.

Second purification step: Equilibration in 10 mM Tris-HCl buffer pH 8.4-1 mM EDTA by passing over a Sephadex G25 resin and then loading onto an ion-exchange resin DEAE-Sephadex equilibrated in 10 mM Tris-HCl buffer pH 8.4, 1 mM EDTA. Washing with 4 volumes of this same buffer and then eluting with a 20 mM KH<sub>2</sub>PO<sub>4</sub> buffer pH 7.4. The pH and ionic strength

Moved to page 51 of the  
substitute specification

MOVED TO PAGE 51  
OF THE SUBMITTED  
MANUSCRIPT

conditions can be modified according to the nature of the hemoglobin variant.

#### Detection of hemoglobin

Hemoglobin (Hb) is detected by virtue of its chromophore, heme, which gives it its characteristic color. At low concentration and in the presence of another chromophore or molecule which scatters light, the signal due to Hb may be masked. This problem can be overcome using a dynamic technique which makes it possible to detect the presence of Hb in a complex system.

This method is based on differential spectra corresponding to a transition between two forms of Hb and on the photodissociation properties of ligands such as O<sub>2</sub> and CO (Gibson, 1956; Mardenet et al., 1994). The probability of dissociation being higher for CO, this ligand is therefore preferably used. The preparation of the samples is carried out under anaerobic conditions.)

(The experimental equipment is composed of two sources of light: the first is a pulsed source (laser) which dissociates the ligands, and the second is a continuous lamp which makes it possible to observe the recombination of the ligands by virtue of a change in the intensity of light transmitted (Figures 4, 5 and 6). The photodissociation is efficient in the entire visible spectral domain; our system consists of a YAG laser whose pulses have a duration of 10 ns at 532 nm. The detection is more sensitive in the Soret band (416 nm); we chose 436 nm close to the maximum absorption of the deoxy form. The changes in transmitted intensity occur first of all in a time of the order of the nanosecond (geminate phase) and then continue in a few milliseconds (bimolecular phase). We are particularly interested in this second phase which reflects the allosteric transitions of Hb (Figures 5 and 6). Rapid and reversible kinetic studies make it possible to obtain numerous data and therefore a )

MOVED TO PAGE 52

END PAGE 51

Moved to  
Page 53  
of the substitute specification

Move to  
Page 53  
of the substitute specification

Move to page 52 of  
The substitute specification

(reliable indication of the state of Hb as regards its normal, physiological function.

The preparation of the samples is carried out as described below. The tobacco leaves (20 g) are ground in liquid nitrogen and then the ground product is mixed with 60 ml of the extraction solution (25 mM Tris-HCl pH 7.5, 10 mM  $\beta$ -mercaptoethanol, 1 mM EDTA). The homogenate is centrifuged at 10000 g at 4°C for 15 minutes. The supernatant containing the soluble proteins is collected. The assay of the proteins is carried out according to the Bradford technique (1976). To 1 ml of plant protein extract (1 mg/ml) are added 32  $\mu$ l and 3.2  $\mu$ l of a concentrated human hemoglobin solution (3.13 mg/ml) in order to obtain solutions containing 100%, 10% and 1% hemoglobin, respectively, relative to the total proteins.

The results obtained are the following:

The kinetics of the samples equilibrated under 0.1 atm CO for three Hb concentrations: 100%, 10% and 1% of the total proteins present in the extract in an amount of 1 mg/ml were measured. The curves are biphasic, similar to those of Hb alone, and exhibit a normal speed (of the order of 1000/s) for the rapid phase (Figure 5). The kinetics are similar for the two concentrations, with the exception of the increase in noise (signal) predictable at low concentration. No signal was observed for the plant extract in the absence of Hb, under the same conditions. We can conclude that the kinetics of recombination of CO with HbA in an extract of tobacco leaves is normal.) end pag 52

#### X: EXTRACTION AND PARTIAL PURIFICATION OF RECOMBINANT HEMOGLOBIN FROM TOBACCO SEEDS

In this section, the techniques used for the detection by Western blotting, the extraction and partial purification and the demonstration of the functionality of the recombinant hemoglobin produced in the seeds of transgenic tobacco plants (rHb), are )

(described. The latter are obtained by transformation of the coexpression plasmid pBIOC 59 containing the cDNAs encoding the  $\alpha$  and  $\beta$  globins allowing targeting in the chloroplast.

a. WESTERN-BLOT DETECTION OF THE RECOMBINANT HEMOGLOBIN ACCUMULATED IN TOBACCO SEEDS

Seventy-five milligrams of tobacco seeds (fresh weight) are ground in liquid nitrogen and then in 600  $\mu$ l of 25 mM ice-cold Tris-HCl buffer pH 8 supplemented with 1 mM EDTA, 1 mM DTT and 1 mM PMSF. The ground product is transferred into an Eppendorf tube and centrifuged at 4°C at 10000 g for 10 min. The supernatant is then concentrated by ultrafiltration with the aid of the micropure .45 and microcon 10 devices (Amicon). The assay of the proteins is carried out according to the Bradford technique (1976) using bovine serum albumin (fraction V) as standard.

The proteins are separated according to their apparent molecular mass by polyacrylamide gel electrophoresis in the presence of SDS according to the Laemmli method (Laemmli, 1970) under reducing conditions. The apparatus used is the Mini-protean II (Bio-Rad). The gel consists of a concentration gel (5% acrylamide, 0.17% bis-acrylamide, 63 mM Tris-HCl pH 6.8, 0.1% SDS) and a separating gel (17% acrylamide) 0.56% bis-acrylamide, 375 mM Tris-HCl pH 8.8, 0.1% SDS). The protein samples are previously diluted with 0.25 volume of loading solution (200 mM Tris-HCl pH 6.8, 400 mM DTT, 40% glycerol, 8% SDS, 0.2% bromophenol blue), then treated at 100°C for 5 min and finally loaded onto the gel. The electrophoresis is carried out in Tris-glycine-SDS buffer (25 mM Tris, 250 mM glycine, 1% SDS) at 25 mA.) *end page 53*

(After electrophoresis, the proteins are transferred onto a nitrocellulose membrane (BA 85, Schleicher & Schuell) by electrotransfer according to the Towbin et al. technique (1979). The transfer is )

MOVED TO PAGE 53

MOVED TO PAGE 53

Moved to page 54 of the  
substrate specification

-88-

(carried out with the aid of the "mini trans blot module" apparatus (Bio-Rad) at 150 V for 90 min in the presence of the transfer solution (25 mM Tris, 192 mM glycine, 20% methanol). The membrane is rinsed for 5 min at room temperature in 1X PBS (10.4 mM Na<sub>2</sub>HPO<sub>4</sub>, 3.2 mM KH<sub>2</sub>PO<sub>4</sub>, 116 mM NaCl) buffer and then dried.

The presence of the globin chains on the Western-blots is detected using, as primary antibody, a rabbit anti-human hemoglobin immune serum (ref: H-4890, Sigma) and, as secondary antibody, an anti-rabbit IgG monoclonal antibody coupled to alkaline phosphatase (A-8025, Sigma). The revealing is performed using the chromogenic substrate [5-bromo-4-chloro-3-indoyl phosphate/nitro blue tetrazolium (BCIP/NBT)].

The membrane is incubated, with stirring, for 5 min in a TBST buffer solution (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Tween 20), and then for at least 30 min in the same solution supplemented with 5% skimmed milk powder (Régilait). The latter solution is replaced, 1/5000 of the volume of anti-hemoglobin immune serum is added and the membrane is incubated for at least 2 hours. It is rinsed 3 times 5 min with TBST solution. The incubation with the secondary antibody is carried out for 1 hour with the anti-rabbit IgG monoclonal antibody diluted 1/10000 in the TBST solution. Next, the membrane is again rinsed 3 times. The alkaline phosphatase activity is revealed by incubating the membrane in the revealing solution (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 330 µg/ml BCIP, 165 µg/ml NBT). The reaction is stopped by rinsing with water.

Figure 7 represents the Western-blot analysis of the protein composition of the extracts of seeds of tobacco plants transformed or otherwise with the plasmid pBIOC59. The polyclonal antibody recognizes the two normal adult hemoglobin (HbA) globin chains separated during the SDS-PAGE electrophoresis. It is observed that the protein extract of the seeds of the transgenic plant T26-22 differs from that of the )

(control plant in the presence of two polypeptides whose apparent molecular mass is similar to that of the globin chains of HbA and which are recognized by the antibody.) Furthermore, they appear to be represented in an equimolar manner. It can therefore be said that in the seeds, the transgenes encoding the fusion proteins transit peptide- $\alpha$  globins and transit peptide- $\beta$  globin are expressed; the cleaving of the transit peptide would be correctly performed, such that the  $\alpha$  and  $\beta$  globins accumulate. In the seeds of 11 plants, of the 20 tobacco plants transformed independently with the plasmid pBIOC59, the presence of the globins is detected. Expressed as equivalents of HbA, the maximum level of about 0.05% rHb relative to the total soluble proteins extracted is observed for the plant T26-22. It was possible to assess this by comparative Western-blot analysis of HbA concentration ranges in the protein extract of control plant seeds.

b. EXTRACTION AND PARTIAL PURIFICATION OF RECOMBINANT HEMOGLOBIN FROM TOBACCO SEEDS.

The partial purification was carried out using as starting material a mixture of the seeds of transgenic tobacco plants transformed with the plasmid pBIOC59 and expressing the rHb.

Fifteen grams of tobacco seeds (fresh weight) are ground in liquid nitrogen and then in 100 ml of 25 mM ice-cold Tris-HCl buffer pH 8 supplemented with 1 mM EDTA, 1 mM DTT and 1 mM PMSF. The ground product is filtered on miracloth® and then the filtrate is centrifuged at 4°C at 10000 g for 10 min. The supernatant is first saturated with carbon monoxide (CO) and then filtered with a 0.22  $\mu$ m filter and finally concentrated by ultrafiltration with the aid of centriprep 10 devices (Amicon). The concentrate is saturated with CO. Two successive chromatographic steps are carried out (4°C) while monitoring the absorbance values at 280 nm (proteins) and 415 nm (hemoproteins).

(i) The concentrate is previously filtered with a 0.22 µm filter and then loaded onto a Sephadryl-S100 column (Pharmacia) (2.1 cm x 90 cm) equilibrated with buffer D (9.12 mM Na<sub>2</sub>HPO<sub>4</sub>, 20.88 mM NaH<sub>2</sub>PO<sub>4</sub>, 1 mM DTT, 1 mM EDTA, pH 6.5). The fraction containing the rHb is collected, filtered through a 0.22 µm filter and then saturated with CO and finally concentrated as above. Sixty-five percent of the proteins are removed at this stage. (ii) This concentrate is loaded onto the second column, a fast-flow S-sepharose (Pharmacia) (1.1 cm x 10 cm) equilibrated with buffer D. After washing with 8 volumes of buffer D, an ionic strength gradient is applied (buffer D to buffer D containing 500 mM NaCl). The hemoglobin is eluted as one peak. The fractions containing this peak are combined and the proteins are concentrated as described above. Before and after concentrating, the samples are saturated with CO. This concentrate constitutes the rHb-enriched fraction called FE-rHb. Only 3% of the proteins of the extract now remain in this fraction. To obtain a control for subsequent analyzes, this purification scheme was applied under the same conditions to an extract obtained from 15 g of tobacco seeds not expressing rHb, leading to the production of the fraction called FE-Control.

The presence of the α and β globins in these fractions was tested for using the Western-blot technique under the conditions described in paragraph X.a. The FE-rHb fraction indeed contains rHb, these two polypeptides being detected (Figure 8).

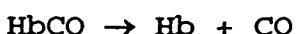
C. DEMONSTRATION OF THE FUNCTIONALITY OF THE RECOMBINANT HEMOGLOBIN BY FLASH PHOTOLYSIS.

The demonstration of the functionality was performed using, as starting material, the rHb-enriched fraction called FE-rHb, using, as control, the equivalent control fraction FE-Control and HbA.)

Moved to page 56 of the  
Sephadryl S-100 Separation

( The control experiments where  $l_F HbA$  was added to the plant extract showed biphasic recombination kinetics and variations of the slow fraction depending on the energy of the flash of laser light. These results demonstrate that the function of HbA is not altered by the solvent conditions used.

After photodissociation of the ligands from Hb, bimolecular recombination occurs within a time scale of  $\mu s-ms$  ( $k_{on}$  speed). Although the natural physiological ligand is oxygen, the studies described were performed with carbon monoxide (CO) which gives a photodissociation signal which is much greater than that obtained with  $O_2$  because the yield is higher. Likewise, the difference in the speeds of recombination for the two conformations of Hb (R and T corresponding to tetramers with and without ligand) is also higher. Experimentally, the samples are equilibrated under 0.1 atm CO which gives the best conditions of observation of the two phases. As) (the reaction is reversible, the photodissociation ( $\gamma$ ) of the same sample can be repeated in order to accumulate several curves, which greatly improves the signal/noise ratio.



The observation of a variation of the amplitude of slow recombination as a function of the dissociation fraction (by modification of the laser energy) demonstrates the presence of a functional hemoglobin.

The transgenic plants receive genetic information only for the synthesis of globin and not for heme. Consequently, if functional Hb (globins + heme) is expressed in the plants, it means that it has captured the heme in situ. Other hemoproteins present in plants can give an optical signal after flash )

(photolysis. These hemoproteins will not give a signal if the heme iron is in the ferric form which does not bind the CO and O<sub>2</sub> ligands. CO and O<sub>2</sub> bind reversibly only if the iron atom is in the ferrous form. It is consequently important to demonstrate the existence of kinetic processes for the two phases and the variation in the relative contributions of the two phases due to factors known to influence the function of hemoglobin.

The enriched sample FE-rHb shows a CO photodissociation signal of 48 mOD (optical density) and makes it possible to carry out certain experiments at different levels of dissociation; these experiments are carried out in the absence of sodium dithionite in order to avoid any parasitic contribution due to the presence of hemoproteins. The same experiment, carried out with the FE-Control fraction showed a signal of 1 mOD (Figure 9).

The results recorded at various levels of laser light energy are shown in Figure 10. The curves are similar to that of HbA and shows the existence of a characteristic property of hemoglobin, namely the lower fraction of slow speed when the light intensity is decreased so as to obtain a lower dissociation.) end page 57

(The sample was then equilibrated under a CO atmosphere. As expected, the recombination kinetics are thereafter more rapid. For hemoglobin in solution, the slow fraction is usually lower at high CO concentration since there is less time available to make the R→T transition after dissociation. The FE-rHb sample does not exhibit this effect (Figure 11).

Another method can be used to study the speeds of association and of dissociation of oxygen. The principle of this method is based on the following fact: although CO has an affinity about 200 times higher than that of oxygen, the speed of association of CO with the ligand-containing Hb (R state) is about 10 times lower than for oxygen. A sample equilibrated with an equal mixture of CO and O<sub>2</sub> will be essentially in the HbCO form. It is then possible to photodissociate the )

Moved to page 57

Moved to page 58

( with a high yield), which allows the study of the recombination of O<sub>2</sub>. A slow terminal phase of the order of 1 s due to the replacement of oxygen by CO provides information on the speed of dissociation (k-off). Only the FE-rHb sample reveals a signal for binding of oxygen (Figure 11).

The studies of the FE-rHb fraction by flash photolysis have shown:

- a biphasic recombination of CO with rapid and slow speeds similar to those observed in tetrameric Hb A;
- a decrease in the slow fraction at low laser energy as for Hb A;
- an increase in speed for higher CO concentrations as for normal Hb;
- a reversible binding of oxygen with on and off speeds similar to those of normal Hb A;

It can be concluded that the recombinant hemoglobin produced in tobacco seeds possesses the properties of tetrameric Hb A in all the functional tests carried out.

**XI: CONSTRUCTION OF CHIMERIC GENES ENCODING THE  $\alpha$  AND  $\beta$  CHAINS OF HUMAN HEMOGLOBIN AND ALLOWING EXPRESSION IN MAIZE SEEDS.) end page 58**

(CONSTRUCTION OF THE PLASMIDS CONTAINING ONE OF THE  $\alpha$  OR  $\beta$  CHAINS OF HUMAN HEMOGLOBIN AND ALLOWING CONSTITUTIVE EXPRESSION OR EXPRESSION IN THE ALBUMIN IN MAIZE SEEDS.

The constitutive or albumin-specific expression, in maize seeds, of the sequences of the  $\alpha$  and  $\beta$  chains of human hemoglobin required the following regulatory sequences:

one of the three promoters allowing a constitutive expression:

- rice actin promoter followed by the rice actin intron (pAR-IAR) contained in the plasmid pAct1-F4 described by McElroy et al. (1991); )

MOVED TO PAGE 58

MOVED TO PAGE 59

Moved to page 59

-09-

( - 35S double constitutive promoter (pd35S) of CaMV (cauliflower mosaic virus). It corresponds to a duplication of the sequences activating transcription, situated upstream of the TATA element of the natural 35S promoter (Kay et al., 1987);

- the promoter of the maize  $\gamma$ zein gene ( $\gamma$ zein) contained in the plasmid py63 (Reina et al., 1990). The plasmid py63 results from the cloning of  $\gamma$ zein at the HindIII and XbaI sites of a plasmid pUC18 containing, between its HindIII and EcoRI sites, the expression cassette "p35S-gus-tNOS" of pBI221 marketed by Clontech. It allows expression in the albumin maize seeds. Combined with the rice actin intron, this promoter confers expression is of constitutive type; one of the two terminators:

- the sequence for termination of transcription, 35S polyA terminator, which corresponds to the noncoding 3' region of the sequence of the circular double-stranded DNA cauliflower mosaic virus producing the 35S transcript (Franck et al., 1980);

- the sequence for termination of transcription, NOS polyA terminator, which corresponds to the noncoding 3' region of the nopaline synthase gene of the Ti plasmid of nopaline-containing strain of Agrobacterium tumefaciens (Depicker et al., 1982).

The type of vector used is derived from pBSIIISK+ (Stratagene). Each vector comprises an expression cassette, namely one of the promoters, one of the  $\alpha$  or  $\beta$  chains of human hemoglobin and one of the terminators. Vectors comprising ~~the two~~ cassettes for expression of each of the  $\alpha$  and  $\beta$  chains of human hemoglobin were also constructed. The clonings were carried out according to the customary methods.)

Moved to  
Page 60

(Bibliographic references:

- Benesch & Kwong. *Hemoglobin* 1994, 18, 185-192.
- Birnboim & Doly. *Nucleic Acids Res.* 1979, 7, 1513-
- Boutry & Chua. *EMBO J.* 1985, 4, 2159-2165.
- Bradford. *Anal. Biochem.* 1976, 72, 248-254.
- Carrer et al. *Mol. Gen. Genet.* 1993, 241, 49-56.
- Chaumont et al. *Plant Mol. Biol.* 1994, 24, 631-641.
- Gamborg et al. *Exp. Cell Res.* 1968, 50, 151-158.
- Guerineau et al. *Nucleic Acid Res.* 1988, 16, 11380.
- Gibson. *J. Physiol.* 1956, 134, 123.
- Edelbaum. *J. Interferon Res.* 1992, 12, 449-453.
- Hanahan. *J. Mol. Biol.* 1983, 166, 557-
- Hanahan. In "DNA cloning volume I, a practical approach" (Ed: Glover D.M.) IRL Press, 1985, pp 109-135.
- Hiatt & Ma. *FEBS Let.* 1992, 307, 71-75.
- Hoffman et al. *Proc. Natl. Acad. Sci. USA* 1990, 87: 8521-8525.
- Horsch et al. *Science* 1985, 227, 1229-1231.
- International Hemoglobin Information Center (1995) *Hemoglobin*, 19, 37-124.
- Jessen et al. *Meth Enzymol.* 1994, 231, 347-364.
- Joshi. *Nucleic Acid Res.* 1987, 15, 6643-6653.
- Kister et al. *J. Biol. Chem.* 1987, 262, 12085-12091.
- Krebbers et al. *Plant Physiol.* 1988, 87, 859-866.
- Marden et al. *Meth Enzymol.* 1994, 232 71-86.
- Mason et al. *Proc. Natl. Acad. Sci. USA* 1992, 89, 11745-11749.
- Moloney. Int. Meeting of Production of Recombinant Proteins in Plants, Leicester 1994, page 36-38
- Murashige & Skoog. *Physiol. Plantarum* 1962, 15, 473-497.)
- (- Nagai & Thogersen. *Meth. Enzymol.* 1987, 153, 461-481.
- Nagai et al. *Proc. Natl. Acad. Sci. USA* 1985, 82, 7252-7255.
- Perutz. *Nature* 1970, 228, 726-739.
- Russel. Int. Meeting of Production of Recombinant Proteins in Plants, Leicester 1994, page 43 )

MOVED TO PAGE 60

MOVED TO  
PAGE 41

Moved to  
page 61

- Sambrook et al. Molecular Cloning: A Laboratory Manual. Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
- Sanger et al. Proc. Natl. Acad. Sci. USA 1977, 74, 5463-5467
- Stephen et al. Nucleic Acid Res. 1990, 18, 7463-7464.
- Svab et al. Proc. Natl. Acad. Sci. USA 1990, 87, 8526-8530.
- Swanson et al. Bio/Technology 1991, 9, 57-61.
- Symons et al. Bio/Technology 1990, 8, 217-221.
- Vanderkerckhove et al. Bio/Technology 1989, 7, 929-932.
- Wagenbach et al. Biotechnology 1991, 9: 57-61.
- Wilson et al. Nucleic Acid Res. 1978, 5, 563-581.
- Kay R. et al., Science, 1987, 236: 1299.
- Franck A. et al., Cell, 1980, 21: 285.
- Depicker A. et al., J. Mol. Appl. Genet., 1982, 1: 561.
- Mc Elroy et al., Mol. Gen. Genet., 1991, 231: 150.
- Reina et al., N.A.R., 1990, 18: 6426.
- Dumoulin et al., Prot. Sci., 1996, 5: 114-120.
- Feng et al., J. Mol. Evol., 1985, 21: 112-115.
- Dumoulin et al., Art. Cells Blood Subs. Immob. Biotech., 1994, 22: 733-738.
- Looker et al., Nature, 1992, 356: 258-260.
- Laemmli, Nature, 1970, 227: 680-685.
- Towbin et al., Proc. Natl. Acad. Sci. USA, 1979, 76: 4350-4354. ) end page 61

CLAIMS

1. Method for producing hemin proteins comprising the following steps:
- i) introducing, into plant cells, one or more nucleic acid molecule(s) each of which comprises at least one sequence encoding a protein component of a hemin protein of animal origin capable of reversibly binding oxygen or a variant or a portion of this protein component, and optionally a sequence encoding a selection agent;
  - ii) selecting the cells containing the nucleic acid encoding the protein component of the hemin protein;
  - iii) optionally propagating the transformed cells, either in culture or by regenerating whole transgenic or chimeric plants;
  - iv) recovering, and optionally purifying, a hemin protein comprising a complex of the protein or proteins encoded by the abovementioned nucleic acid with at least one iron-containing porphyrin nucleus, or a plurality of these complexes.
2. Method according to Claim 1, characterized in that the hemin protein is a single polypeptide chain, for example myoglobin.
3. Method according to Claim 1, characterized in that the hemin protein is a heterooligomer, the nucleic acid(s) comprising the sequences encoding each of the various protein units.
4. Method according to Claim 3, characterized in that the hemin protein is human hemoglobin, or a derivative thereof, the nucleic acid(s) comprising sequences encoding  $\alpha$  and  $\beta$  globin, or variants of  $\alpha$  or  $\beta$  globin, the variants differing from the natural sequence in one or more amino acid substitution(s), deletion(s) or insertion(s).
5. Method according to Claim 4, characterized in that the sequences encoding the various protein units,  $\alpha$  and  $\beta$  globin, are contained within the same nucleic acid molecule. }

Moved to page 62

move to page 2

- 73 -

(5. Method according to Claim 4, characterized in that the sequences encoding the various protein units,  $\alpha$  and  $\beta$  globin, are contained within separate nucleic acid molecules.)

(7. Method according to any one of Claims 1 to 6, characterized in that the introduction of the nucleic acid molecule(s) is carried out by transformation of the nuclear genome of the plant cell.

8. Method according to Claim 7, characterized in that the sequence encoding the protein component comprises one or more sequence(s) encoding chloroplast targeting signals, or mitochondrial targeting signals.

9. Method according to Claim 7, characterized in that the sequence encoding the protein component comprises one or more sequence(s) encoding an N-terminal signal peptide and optionally a signal responsible for retaining the protein in the endoplasmic reticulum, or a vacuolar targeting signal.

10. Method according to any one of Claims 1 to 6, characterized in that the introduction of the nucleic acid is performed by transformation of the mitochondrial or chloroplast genome.

11. Method according to any one of Claims 1 to 9, characterized in that the nucleic acid comprises, in addition to the coding sequence(s), sequences for regulation of transcription which are recognized by plant cells.

12. Method according to any one of Claims 4 to 11, characterized in that the coding sequence(s) encode(s) a hybrid molecule composed of at least the active parts of  $\alpha$ -globin and  $\beta$ -globin.

13. Method according to any one of Claims 1 to 12, characterized in that it comprises, between the propagation step and the recovery step, a step for detecting hemin proteins, and in particular hemin proteins whose porphyrin nucleus consists of iron-containing protoporphyrin IX.

14. Method according to Claim 13, characterized in that during the step for propagating plant cells, iron- )

move to page 63

MOVED TO  
PAGE 63

(containing protoporphyrin IX is added to the culture medium used for growing the cells.

15. Hemin protein having the capacity to reversibly bind oxygen, characterized in that it comprises at least one iron-containing porphyrin nucleus, of plant origin, and a protein component comprising at least one polypeptide chain, of animal origin.

16. Protein according to Claim 15, characterized in that the iron-containing porphyrin nucleus is iron-containing protoporphyrin IX, or a protoporphyrin differing from protoporphyrin IX in the nature of the side chains carried by the  $\beta$  atoms of the pyrrole rings.

17. Protein according to Claim 16, characterized in that the protein component comprises at least one  $\alpha$ - and/or  $\beta$ -globin polypeptide chain, or variants thereof comprising one or more amino acid substitution(s), deletion(s) or insertion(s), the hemin protein being capable of binding oxygen reversibly.

18. Protein according to Claim 17, characterized in that the  $\alpha$ - or  $\beta$ -globin chain, or variants thereof, comprises in addition a chloroplast targeting signal, a mitochondrial targeting signal, or a N-terminal signal peptide optionally in combination with a signal responsible for retaining protein in the endoplasmic reticulum, or a vacuolar targeting signal.

19. Protein according to Claim 17, characterized in that each polypeptide chain lacks  $NH_2$ -terminal methionine.

20. Protein according to any one of Claims 17 to 19, characterized in that the protein component comprises at least four polypeptide chains of  $\alpha$ - and/or  $\beta$ -globin or variants thereof, each polypeptide chain being bound to an iron-containing protoporphyrin nucleus.

21. Protein according to Claim 20, characterized in that it comprises 2  $\alpha$ -globin chains and 2  $\beta$ -globin chains, or variants thereof. )

22. Protein according to any one of Claims 15 to 21, characterized in that it binds oxygen with an affinity of between 7 and 40 mm Hg, preferably 15 to 20 mm Hg.
23. Nucleic acid comprising:
- i) one or more sequence(s) encoding a protein component of an animal hemin protein, the said protein having the capacity to reversibly bind oxygen, and
  - ii) sequences for regulation of transcription which are recognized by a plant cell, comprising a promoter and sequences for regulation of termination, and
  - iii) one or more sequence(s) encoding a targeting signal of plant origin.
24. Nucleic acid according to Claim 23, characterized in that the regulatory sequences comprise one or more promoter(s) of plant origin.)
25. Nucleic acid according to Claim 23 or 25, characterized in that the sequences encoding the targeting signal encode a mitochondrial or chloroplast targeting peptide called "transit" peptide.
26. Nucleic acid according to Claim 23 or 24, characterized in that the sequences encoding the targeting signal encode an N-terminal signal peptide of plant origin, optionally in combination with a sequence encoding an endoplasmic retention signal or a vacuolar targeting signal.
27. Nucleic acid according to any one of Claims 23 to 26, characterized in that the coding sequence encodes human  $\alpha$ - or  $\beta$ -globin, or a variant thereof differing from the natural sequence in one or more amino acid substitution(s), deletion(s) or replacement(s), or a portion of human  $\alpha$ - and/or  $\beta$ -globin.
28. Nucleic acid according to any one of Claims 23 to 27, comprising, in addition, one or more intron(s), preferably of plant origin.
29. Nucleic acid according to any one of Claims 23 to 28, characterized in that the sequence encoding the protein component is a cDNA.

Moved to page 65

Moved to page 64

- ( 31. Vector comprising one or more nucleic acid molecule(s) according to any one of Claims 23 to 29.
31. Plant cells transformed in a stable manner by the nucleic acid according to any one of Claims 23 to 29.
32. Plant cells capable of producing one or more hemin protein(s) according to any one of Claims 15 to 22.
33. Plant cells according to Claim 32, characterized in that they comprise nucleic acid comprising one or more sequence(s) encoding a protein component of the said hemin protein in association with one or more sequence(s) for regulation of transcription recognized by the cell.
34. Plant cells according to any one of Claims 31 to 33, characterized in that they are a culture of plant cells, for example in liquid medium or immobilized cells, or a root culture.
35. Plant cells according to any one of Claims 31 to 33, characterized in that they are cells which form part of a whole transformed plant.)
- ( 36. Chimeric or transgenic plant capable of producing one or more hemin protein(s), for example hemoglobin or a derivative thereof, characterized in that it comprises cells according to any one of Claims 31 to 33.
37. Seeds of transgenic plant according to Claim 36.
38. Pharmaceutical product comprising one or more hemin protein(s) according to any one of Claims 15 to 22 in association with a physiologically acceptable excipient.
39. Hemin proteins according to any one of Claims 15 to 22 for use as medicament.
40. Hemin protein according to Claim 39, for use in the treatment of conditions requiring an improvement in the transport of oxygen in the blood.
41. Use of a hemin protein according to any one of Claims 15 to 22 for the preparation of a medicament for )
- Moved to page 65

- / / -

(the treatment of conditions requiring an improvement in  
the transport of oxygen in the blood.

42. Use of a hemin protein according to any one of  
Claims 15 to 22 in an industrial, cosmetic product or  
as chemical reagent. )

*Moved  
to  
page 66*

## C ABSTRACT

A method for producing haemin proteins by (i) inserting into plant cells one or more nucleic acid molecules that each comprise at least one sequence coding for a protein component of an animal haemin protein capable of reversibly binding oxygen, or for a variant or portion of said protein component, and optionally a sequence coding for a selection agent; (ii) selecting cells containing nucleic acid coding for the protein component of the haemin protein; (iii) optionally propagating the transformed cells either in a culture or by regenerating whole transgenic or chimeric plants; and (iv) recovering and optionally purifying a haemin protein that includes a complex consisting of the protein or proteins coded for by said nucleic acid and at least one iron-containing porphyrinic nucleus, or a plurality of such complexes. )

MOVED  
to Page 67